

EXHIBIT D

Variety-Based Congestion in Online Markets: Evidence from Mobile Apps

Daniel Ershov*
UCL School of Management
d.ershov@ucl.ac.uk

August 10, 2022

Abstract

In many online markets, consumers have to spend time and effort browsing through products. The addition of new products could make other products less visible, creating congestion externalities. Using Android app store data, I take advantage of a natural experiment – a re-design of part of the store – to show evidence of congestion externalities online: more apps in the market directly reduce per-app usage/ downloads. The natural experiment also increases long-run entry, but a structural demand model that accounts for congestion externalities suggests that forty percent of consumer variety welfare gains are lost from higher congestion.

*I am grateful to the editor, Leslie Marx, and three anonymous referees for their helpful comments. This paper was a chapter in my PhD dissertation and previously circulated under the titles “The Effects of Consumer Search Costs on Entry and Quality in the Mobile App Market,” and “Consumer Product Discovery Costs, Entry, Quality and Congestion in Online Markets.” I would like to thank Victor Aguirregabiria, Avi Goldfarb and Heski Bar-Isaac. This paper also substantially benefitted from discussions with Joel Waldfogel, Eduardo Souza Rodrigues, Mar Reguant, Sandro Shelegia, and comments from participants at numerous seminars and conferences. I gratefully acknowledge past support received from SSHRC, the Ontario Government, the University of Toronto and ANR under grant ANR-17-EUR-0010 (Investissements d’Avenir program). All errors are my own.

1 Introduction

In the last 30 years, digital-technology related reductions in production, distribution, and transportation costs facilitated the creation of many new products. Information-aggregating online marketplaces / platforms unlocked consumer access to these products. However, actually discovering new products online remains a challenge for consumers. Despite platform features such as personalized recommendations, search algorithms, product categories, bestseller lists, and ratings, consumers have to spend substantial time and effort examining products. Limited attention and screen size creates congestion, as consumers may not see some products because of the existence of others. Such congestion is a source of consumer and firm concerns in virtually every online market ([TheAtlantic.com](https://www.theatlantic.com)), and could undercut the benefits of growing variety.

This paper presents the first empirical evidence on the existence and magnitude of congestion effects online. I use data on the Google Play (Android) mobile app store from January 2012 to December 2014. This is an online market with over 100 million US consumers and mostly free products/apps. The Google Play Store is representative of many online markets. Thousands of new apps appear every week and discoverability is a major concern ([iMediaConnection.com](https://www.imediaconnection.com)). Consumers primarily discover products by browsing through platform-defined categories (e.g., “Productivity Apps”).¹ Many other online markets such as eBay, Amazon, and Netflix are similarly organized.

To establish the existence of congestion externalities, I take advantage of a re-design of part of the Google Play Store. App stores are split into “game” and “non-game” sections. In March 2014, Google Play re-organized its 6 game categories into 18 categories (Table 1), reducing the number of apps per category.² As non-game categories were not changed, I compare game and non-game outcomes using a difference-in-differences approach. I show that, in the very short-run, downloads increased for games relative to non-games, and most importantly, downloads increased by more for apps that ended up in less populous categories. This points to the presence of congestion externalities, as apps in categories with fewer other apps are more likely to be visible to consumers.

The re-design also had an effect on the supply of game apps, increasing game entry by approximately 34%. Although greater variety should increase welfare, the existence of congestion externalities should mitigate welfare gains from entry. Drawing on a consideration-set framework, I estimate the parameters of an app demand

¹See Section 2 and Online Appendix A.3.

²The 24 non-game categories do not change and are listed in Table A1 in Online Appendix A.

model that accounts for congestion externalities. I then evaluate the welfare effects of congestion. Each consumer gained \$0.057 per-month from the additional product variety in the market after re-categorization. This adds up to nearly \$70 million per-year across all US Google Play consumers. However, congestion externalities dissipate approximately 40% of these gains.

Table 1: **Google Play Game Categories Before and After March 2014**

Before :

Arcade & Action, Brain & Puzzle, Card & Casino, Casual, Racing, Sports

After :

Action, Adventure, Arcade, Board, Card, Casino, Casual, Education, Family, Music, Puzzle, Racing, Role Playing, Simulation, Sports, Strategy, Trivia, Word

This paper’s main contribution is to show the economic importance of congestion externalities online. Previous literature highlighting the benefits of product variety coming from online markets and digitization, such as [Brynjolfsson et al. \(2003\)](#), and [Aguilar and Waldfogel \(2018\)](#), did not capture this limiting factor. [Quan and Williams \(2018\)](#) also identify the limited benefits consumers receive from additional product variety, although there, the main friction comes from heterogeneity in consumer preferences, rather than congestion affecting consumer product discovery.

This paper also shows that changes in platform design can increase product variety on the platform.³ It confirms theoretical predictions about platform design and entry ([Anderson and Renault \(1999\)](#), [Bar-Isaac et al. \(2012\)](#), [Goldmanis et al. \(2010\)](#)), and descriptive evidence ([Brynjolfsson et al. \(2011\)](#), [Zentner et al. \(2013\)](#)).

Product assortment, as measured by the number of products, is a key competitive outcome of concern for regulators in most online markets.⁴ A key novel trade-off coming out of my results is that platform-led or regulator-led policy that encourages entry benefits consumers, but also intensifies congestion externalities. Personalization and recommendation technology has improved since the time period studied in this paper. Many online markets (e.g., Amazon, video streaming services) have better such technology today than the mobile app market in 2012-2014. The effects generated by a platform’s category re-design would likely be smaller today and in

³Previous empirical literature primarily focuses on the effects of platform design on price competition and match quality - e.g., [Fradkin \(2017\)](#) and [Dinerstein et al. \(2018\)](#).

⁴In many markets prices are zero or uniform, so changes in competition intensity, platform design, or other market conditions cannot have an effect on prices.

the future. Nonetheless, my findings serve as a useful upper bound on congestion effects that could be generated by future platform-led or regulator-led policy aiming to increase entry online.

The paper proceeds as follows: Section 2 provides an overview of mobile app market and the re-categorization event. Section 3 describes the data and presents some summary statistics. Section 4 presents the reduced form evidence. Section 5 discusses the specification and estimation of the structural model and the welfare decomposition. The final section concludes.

2 Background

2.1 Consumer Browsing on the Google Play Store

When a consumer opens the Google Play Store on their phone circa 2014, their first choice is between games or non-game apps. After this, the consumer chooses a category to look at more specific game or non-game product types. Once they choose a category, they have several lists of apps to choose from. Bestseller-lists display apps with the largest number of downloads over a certain period of time.⁵ Other lists are *not* organized based on past popularity, but showcase “featured” apps selected by human editors or algorithmically. There are also lists uniquely dedicated to showcasing new apps. Unlike bestseller-lists, which often feature the same apps for long periods of time, apps in other lists within a category rotate frequently. During my sample period, consumers *cannot* filter (or selectively browse) categories by app ratings, number of downloads, or other app characteristics. As well, during my sample period, the search function in the Google Play Store was unreliable (see additional discussion in Section 2.2). Before downloading an app, consumers observe a number of screenshots from the app, its average rating, how many people have downloaded this app, the size of the app in MB, and a text describing the app.

Consumers have a search bar at the top of the screen, and can use it to find apps. They also receive some personalized recommendations on the home page of the Google Play Store. However, surveys of app consumers during and after my sample period also show that browsing the category structure, as opposed to using a search engine or targeted recommendations, is the primary way in which consumers discover new products.⁶ The search function in the Google Play Store was unreliable

⁵The exact algorithm determining the position of an app in the top lists is unknown, but it is related to downloads (Adweek.com).

⁶58% of Android consumers discover new products through “General browsing in the app store”, according to a Forrester survey from 2012. A Nielsen survey from the same year, and other surveys

([AndroidAuthority.com](#)). Google introduced app indexing in Google Search only in early 2015 (after my sample period) but this functionality was not widely adopted by app developers. It required substantial code adjustments and also frequently failed to provide working links to the Google Play Store ([SmashingMagazine.com](#)). The “personalized” recommendations primarily featured already popular apps that consumers likely already downloaded.

2.2 March 2014 Game Re-Categorization Event

From 2009 to 2014, the Google Play Store had six game categories, compared to eighteen game categories in the Apple iOS store. Anecdotal reports suggested one reason for this discrepancy was Google’s smaller initial app variety in 2009. Google also expected consumers to primarily use the search bar within the store or for Android apps to be quickly integrated into Google Search results. As described above, this was not the case by 2014.

On December 9, 2013, Google announced that it was expanding the number of game categories in the first quarter of 2014 ([droid-life.com](#)). The new classification matched Apple’s existing eighteen category structure. Developers with existing apps could choose their own new categories ahead of time. The date for the re-categorization was announced to be February 2014, but Google eventually delayed launching the store re-design until March 17, 2014. Industry observers and developers were reportedly surprised both by the initial announcement and by launch delays ([AndroidPolice.com](#)).

Since the re-categorization tripled the number of game categories, it immediately reduced the number of apps in each category. Categories are important to the consumer search process, and the number of apps per category is important to an app’s visibility. Tripling the number of categories effectively reduces competition for scarce visible spaces in the store.⁷ An app is more likely to feature in the rotating “featured” app lists and be visible to consumers.⁸

from 2016 also have similar findings. See Figures [A2](#) and [A3](#) for more.

⁷Given the small number of categories, choosing a category should have negligible cost for consumers. However, if the number of categories is large, the categories themselves could compete for scarce consumer attention, and increasing the number of categories could also make it harder for consumers to find products, for the same reason that increasing the number of apps does. That is not the case in this particular market. See additional discussions in Sections [4.2](#) and [5.4](#).

⁸The titles of categories also became more informative about the types of apps present. Before the new categories, consumers looking for music, family, or strategy games did not know precisely where to look. After re-categorization, this changed. I show evidence for this in Online Appendix [C.2](#). On the supply side, developers knew that if they produce such a game, there is a clear place

These changes were consistent with Google’s primary motivation for the re-design: improving the consumer search experience. An industry observer described that “searching for general apps on the Play Store is an exercise in frustration” under the old six game category system ([AndroidPolice.com](#)). In the PR announcement for the 2014 change, Google stated that the new game categories “mak[e] it easier for players to find games they’ll love” ([AndroidPolice.com](#)). A Google blog post discussing a subsequent re-categorization in 2016 clearly stated that new app categories “improve the overall search experience” and “mak[e] them more comprehensive and relevant to what users are looking for today” ([GoogleBlog.com](#)). For the 2014 game re-categorization studied here, the wholesale adoption of Apple’s game categories suggests that Google did not choose categories based on pre-existing trends.⁹

The 2014 re-design appears to have been successful for Google, as evidenced by subsequent category expansions. In 2015 Google introduced additional “Family Game” sub-categories to make it easier for parents to find games for kids of varying ages ([AndroidCommunity.com](#)). Google again introduced new categories in 2016 for non-game apps. ([GoogleBlog.com](#)). The Apple App store also changed their categories, experimenting with removing some non-game app categories in 2018 ([MacObserver.com](#)). These events fall outside the scope of my data. Re-designs also happen in other online markets. Amazon changes its product categories frequently, and eBay also experimented with re-design (see [Dinerstein et al. 2018](#)).

3 Data

3.1 Data Description

My data comes from AppMonsta.com and consists of daily snapshots of all apps on the US Google Play store, aggregated at the weekly or monthly level, starting from January, 2012 and up to December, 2014.¹⁰ This is the first paper to use this dataset.¹¹ The data contains all information that consumers observe on the Play store - app price (in USD), a histogram of the ratings the app received (ranging from

for it to be discovered. This is especially the case since the new categorization structure already existed on the iOS Store for years at that point, and developers frequently multi-homed.

⁹Formal tests for downloads in Figure C1 and Online Appendix C.4.1, and for entry in Figure C3 show sharp changes in outcomes exactly around the period of re-categorization but no differences before the policy’s announcement. Private discussions with Google employees familiar with changes to the Play Store also confirm that app entry is not a consideration when making such changes. They suggest opposite concerns about too many products already appearing in the store.

¹⁰Weekly aggregation is only used to predict app downloads (see Online Appendix B.2).

¹¹[Liu et al. \(2014\)](#) use an app dataset from the same provider for 2011-2012.

1 to 5), app size (in MB), the number of preview screenshots the app shows, the number of video previews the app shows, and a download range for the app (number of lifetime downloads). I also observe the app’s category, the name of the app’s developer, and a text description of the app.

I supplement this data with scraped historical app rankings from Flurry.com. Using this data, I observe the “top lists” for every category in each week, which approximate the top 500 weekly best-selling free and paid apps in each category.¹²

3.2 Data Management

3.2.1 App-Type Classification

I classify game apps into *types* based on their categories in the last period of the sample.¹³ An app belonging to the Music category in the last period after re-categorization is defined to be a “Music” type game.¹⁴ Table B4 in the Online Appendix shows some summary statistics at the app-type level. There are 42 app-types, 18 of which are game types. Game app-types have fewer apps than non-game types. The average game type is less than a quarter of the size of the average non-game type.

Game app-types, denoted by c for the rest of the paper, are distinct from categories, denoted by c^* . Before re-categorization, multiple app-types can be present in one category. Before re-categorization, apps could also be present in categories whose names do not represent their type. For example, in 2012, Music or Family type games can be in the Arcade & Action category, or in the Brain & Puzzle category. For non-games, the distinction between categories and app-types does not practically matter, as fewer than 0.2% of apps change categories in my sample.¹⁵

¹²See additional discussion in Online Appendix B.2.

¹³Apps can also be classified into a larger number of types based on alternative criteria, as in Kesler et al. (2020). However, it is likely that app developers think about their apps in terms of the 18 post re-categorization types because the Apple App Store already had that category structure for years and developers often multi-home.

¹⁴There are two possible drawbacks to this approach: (i) Approximately 1% of games exit the market before re-categorization and cannot be classified in this way. I use app descriptions to classify *only* these apps into types. See Online Appendix B.1 for a description of this approach. Robustness checks show that dropping these apps does not change the baseline results. (ii) There is possibly some selection of apps into categories based on competition and other features of the market. i.e., it is possible that a “Music” type app enters into another category. However, both of these concerns are likely minor in practice. Previous versions of the paper used a machine learning algorithm and text-based analysis to classify *all* apps into types. Results are quantitatively similar across the two classification approaches.

¹⁵This is likely because developers are very careful in deciding on the initial category positioning.

3.2.2 Predicted Downloads

In the raw data, I do not observe apps’ weekly or monthly downloads, but only life-time download bandwidths reflecting how many downloads an app has had throughout its entire time in the store.¹⁶ These can range to millions of downloads.¹⁷ I estimate monthly downloads using information about app rankings in each category and week, and the download bandwidth of new apps (apps that arrived in the market in that week). I recover a relationship between rankings and downloads for new apps.¹⁸ Then, I predict the downloads of all other apps in the market. Section B.2 in the Online Appendix provides more details about this procedure.¹⁹

This approach relies on a functional form assumption for the relationship between downloads and rankings, and could produce inaccurate estimates of monthly downloads (Liebowitz and Zentner 2020). As a robustness check, I use an alternative proxy for downloads: the difference in the number of ratings an app receives between two consecutive months. Results are qualitatively similar across the two approaches (see Online Appendix C.4.3).

3.3 Descriptive Statistics

Table 2 shows some summary statistics at the app level. There are approximately 33.7 million app-month observations in total, consisting of 2.6 million unique apps. Of these, 17% belong to game categories. 20% of apps have non-zero prices. The average price for a paid app is approximately \$3.3. I provide additional summary statistics for the sub-sample of game apps I use in Section 5 to estimate the demand model. Average price for paid games is approximately \$1.9.

Re-positioning is risky, as it moves an app to compete with a new set of other apps for scarce consumer attention and could push it down a bestseller list or make the app less likely to be “featured” on the store otherwise.

¹⁶These download measures do not include updates, and they also do not double count the downloads of different versions of the app.

¹⁷Table B1 in Online Appendix B.2 shows all download bandwidths.

¹⁸Intuitively, I observe a new app ranked 1st in a category with a lower bound of 50,000 downloads, a new app ranked 10th with a lower bound of 10,000 downloads, and a new app ranked 100th with a lower bound of 500 downloads. Under certain distributional assumptions, I can recover the relationship between the lower bound of downloads and ranking. See additional discussion in Online Appendix B.2.

¹⁹I predict zero downloads for about 20% of apps in a given period. To mitigate the “zeroes” problem in demand estimation, under some assumptions about the underlying distribution of consumer downloads I apply the method of Gandhi et al. (2014).

Table 2: **App Summary Statistics**

Variable	Mean	Std. Dev.	N
<i>App Level</i>			
Game App Dummy	0.168	0.374	2.6 million
Paid App Dummy	0.2	0.4	2.6 million
<i>App-Month Level</i>			
Lifetime Downloads (Min.)	38,261	1.9 million	33.7 million
App Size (in MB)	21.99	29.75	33.7 million
Monthly Predicted Downloads	559	25,169	33.7 million
Number of Screenshots	4.71	3.54	33.7 million
Number of Videos	0.09	0.28	33.7 million
Mean Rating	4.0	0.66	33.7 million
Price (for Paid Apps)	3.27	8.93	6.8 million
<i>App-Month Level for Section 5 Sample: Game-Apps</i>			
Lifetime Downloads (Min.)	64,269	904,779	4,152,147
App Size (in MB)	11.645	42.982	4,152,147
Monthly Predicted Downloads	236	3,509	4,152,147
Number of Screenshots	5.827	4.564	4,152,147
Number of Videos	.176	.38	4,152,147
Mean Rating	3.355	1.618	4,152,147
Price (for Paid Apps)	1.92	4.754	796,522

4 Reduced Form Evidence of Congestion

4.1 Main Results

In this section, I use the March 2014 re-categorization event to test whether app store congestion - driven by the number of apps - has an effect on consumer demand and usage. My main demand outcomes are aggregate and app-level downloads, since I do not have direct data on app usage.²⁰ I have two identification strategies to test for the presence of congestion externalities.

²⁰Downloads may be more correlated with usage for paid apps than for free apps. I estimate effects separately for paid apps online in Online Appendix C.4.2. Results are qualitatively and quantitatively similar to those in the main text, suggesting that in this setting, downloads closely correlate with usage.

My first strategy uses a heterogeneous treatment effects difference-in-differences approach. Store re-design affected game apps, but not non-game apps. The two groups of apps were separated in the store, and the timing of the policy was a surprise (see Section 2).²¹ I focus on a subset of game app-types.²² Four post-policy game categories are derived from two pre-policy categories: Arcade & Action split into Arcade and Action games, and Card & Casino split into Card and Casino games.²³

Although they originate from the same pre-policy category, the number of apps between the split app-types was unequal before re-categorization. There were three times as many Arcade games as Action games, and three times as many Card games as Casino games. This means that before re-categorization, a Card game and a Casino game had the same number of other apps in their category, affecting their chances of being visible to consumers. After re-categorization, Card games had more other apps in their category than Casino games. If congestion exists and the number of apps in a category matters for discovery, re-categorization should immediately benefit Action games more than Arcade games, and Casino games more than Card games.

I test whether immediate changes in downloads after re-categorization were larger for the app-types with fewer apps by estimating the following regression for app-type c (or app j) at time t :

$$y_{(j)ct} = \tau^1 \text{Game}_c \times \text{Post}_t + \tau^2 \text{Game}_c \times \text{Post}_t \times \text{Small Type}_c + \delta_t + \delta_{(j)c} + e_{(j)ct} \quad (1)$$

where δ_t are time fixed effects, $\delta_{(j)c}$ are app-type (or individual app) fixed effects, Post_t is a dummy equal to one after re-categorization from March 2014, and Game_c is a dummy variable equal to one for the game types/apps and zero for non-game types/apps. Small Type_c is a dummy equal to one for Action and Casino apps/app-types, and zero otherwise. The key parameter in this regression is τ^2 , identifying the average heterogeneity in treatment effects between Arcade and Action, and Cards and Casino games. The absolute baseline throughout is the non-treated group (non-game apps).

²¹Formal tests for parallel trends for downloads are in Online Appendix C.4.1.

²²Average download effect estimates on all game categories are in Table C1 in the Appendix.

²³Ten app-types did not have pre-existing categories: Adventure, Board, Education, Family, Music, Role Playing, Simulation, Strategy, Trivia, and Word games. These apps were also affected by the policy through a different channel - improved informativeness of the search process. I discuss it more in Appendix C.2. Of the remaining game types that existed as categories before re-categorization, Casual, Racing, and Sports games did not formally change. Brain & Puzzle transformed into Puzzle games.

I estimate the regression both at the app-type and the individual app level. The sample includes the four game app-types described above and all non-game app-types. The sample period is the four months, from January 2014 to April 2014, to limit confounding changes in product assortment.²⁴

Estimates from this regression, at the app-type and app levels, are in Columns (1) and (2) in Table 3. They show that downloads increase more after re-categorization for smaller sub-categories/types as compared to larger sub-categories/types. This heterogeneity is substantial. The baseline average increase in downloads for large types is 20-30%, but downloads increase for small sub-categories by 100-120%. At the app-level, these effects are conditional on app fixed effects and additional app controls. This evidence is consistent with congestion externalities, as the actual number of Card and Casino games is not changing much between just before and just after re-categorization. What changes is the number of irrelevant alternatives that compete for scarce visible space, which decreases more for Casino than for Card games and more for Action than Arcade games.

My second identification strategy takes advantage of the fact that re-categorization reduced the number of apps per category for *all* game categories, and not just for categories that were split. Even a category such as Racing games, whose title was not changed, saw reductions in the number of apps. This was likely because some pre-policy “Racing” apps were better described as Music or Family games, and fit better into one of the new category spaces. As a result, for any game app, the number of other apps in their category declined immediately after re-categorization.²⁵ As before, the differences in the number of other apps in its category immediately after re-categorization, as compared to before re-categorization, are not driven by entry.

I use this to provide further evidence relating changes in congestion through the number of apps in a category to downloads. For each *game j*, I regress the first difference in downloads on the first difference in the number of apps in their category around the period of re-categorization.²⁶ Differencing absorbs most time-invariant app characteristics, and I also control for other app characteristics like average app

²⁴I focus on four months rather than just February and March. Since the re-categorization happened roughly in the middle of March, effects identified using February and March are likely incomplete. Results using only February and March 2014 are in Table C3. They are qualitatively identical, but quantitatively smaller than in Table 3. Null estimates for non-existent (“placebo”) events before and after actual re-categorization are in Table C4.

²⁵See Figure C2 in the Online Appendix for direct evidence of this.

²⁶I also estimate the same regression with the difference in paid app prices as the dependent variable. This regression helps verify that the change in the number of apps in a category operates primarily as a change in congestion, rather than a change in competition intensity. See Online Appendix C.8 for additional discussion.

Table 3: Reduced form Congestion Evidence

<i>Outcome Variable:</i>	$\ln(\text{Tot. Type Downloads})$ (1)	$\ln(\text{App Downloads})$ (2)	Post/Pre $\Delta \ln(\text{App Downloads})$ (3)
Games \times Post	0.339* (0.129)	0.229 (0.111)	
Games \times Post \times Small Type	0.745*** (0.085)	1.028*** (0.153)	
Post/Pre $\Delta \ln(N \text{ Apps in Category})$			-0.651*** (0.003)
Unit of Observation:	App-Type	App	App
Sample Period:	Jan 14/Apr 14	Jan 14/Apr 14	Jan 14/Apr 14
Sample:	All Non-Games + Arcade, Action, Card and Casino	All Non-Games + Arcade, Action, Card and Casino	All Games
Year/Month FE	•	•	
App-Type FE	•		
App FE		•	
App Controls		•	•
Observations	112	4,770,936	142,254
R-squared	0.953	0.984	0.879

Notes: The sample period in all columns is January 2014-April 2014. Data in Columns (1) and (3) consists of monthly observations at the app-type level. Data in Columns (2) and (3) consists of monthly observations at the app level. Columns (1) and (2) include all non-game apps and Arcade, Action, Card and Casino games. Column (3) includes all game apps. Outcomes for Columns (1) and (2) are the natural logarithms of downloads at each aggregation level. The outcome for Column (3) is the difference between the natural log of average app downloads in March and April 2014 and the natural log of average app downloads in January and February 2014. Controls include year and month fixed effects and app-type or app fixed effects depending on the column. Column (3) does not include app fixed effects, since observations are already in first differences. Additional app-level controls include average app ratings, a dummy for whether an app is free or paid, the price of an app if it is paid and app-age specific fixed effects. The variable “Games \times Post” is a dummy variable equal to 1 for games (or game app-types for even columns) during and after March 2014. The variable “Games \times Post \times Small Type” is a dummy variable equal to 1 during and after March 2014 only for Action and Casino games. $\Delta \ln(N \text{ Apps in Category})$ is the difference in the natural log of the number of apps in the category of app j after re-categorization and the natural log of the number of apps in the category of app j before re-categorization. Standard errors are clustered at the app-type level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

rating and app price. The estimating equation is:

$$\ln(\text{Downloads}_{j,Post}) - \ln(\text{Downloads}_{j,Pre}) = \alpha(\ln(N\text{Apps})_{jc,Post}) - \ln(N\text{Apps})_{jc^*,Pre}) + \beta X_j + \epsilon_j \quad (2)$$

where $\text{Downloads}_{j,Post}$ are game j 's average monthly downloads in the two months after re-categorization, and $\text{Downloads}_{j,Pre}$ are game j 's average monthly downloads in the two months before re-categorization.²⁷ $N\text{Apps}_{jc,Post}$ and $N\text{Apps}_{jc^*,Pre}$ are sim-

²⁷In Table C3, I show that the results hold using only February and March data.

ilarly the number of apps in the category of app j before and after re-categorization.²⁸

Estimates of this regression are in Column (3) of Table 3. They show a statistically and economically significant negative relationship between the number of apps in a category and app downloads. A one percent increase in the number of apps in a category reduces app downloads by 0.65 percent.²⁹ The average growth rate of the number of apps-per-category in game categories after re-categorization was 9%, suggesting that congestion can have substantial effects on consumer demand. In Online Appendix C.6 I find a similar elasticity using longer-run variation in the number of apps in a category.

4.2 Discussion

The estimates above suggest the existence of variety-driven congestion externalities in the mobile app market. This evidence matches industry analyst statements about the importance of creating products in niches with relatively fewer competitors and maximizing chances of being featured on the store ([PrioriData.com](https://prioridata.com)). It also reflects consumer complaints about the existence of too many apps in the market, and the difficulty of finding apps ([Forbes](https://www.forbes.com)). The presence of congestion raises questions about the welfare gains consumers experience from additional product variety. When variety increases in a market, despite the *existence* of additional products, consumers only have access to a limited subset of these. I quantitatively evaluate the importance of congestion effects on consumer welfare in Section 5.

The evidence presented above comes from a single natural experiment in a particular market, which introduces concerns about the generality of the results. Would the effects be different in a market with better personalization, or better search-bar technology? Would the effects be different if there were more categories?

A similar platform re-design that triples the number of categories would likely produce different effects in a market with hundreds or thousands of categories, such as Amazon. In a market with many categories, more consumer time and effort would be spent browsing *between* categories, as compared to *within* categories. In an extreme example, a market with a category for each product is analogous to a

²⁸Before re-categorization app j 's category is represented by c^* . After re-categorization app j 's category coincides with its type, c .

²⁹This may be a biased estimate of the true effect, as apps can strategically sort into categories. However, if that is the case, the main concern comes from apps with better demand shocks or unobservables sorting into “better” categories with fewer other apps. This would suggest a *negative* correlation between the error term and the number of apps in a category, and a *positive* correlation between the error term and downloads. Together, this means I under-estimate (in absolute terms) the effects of the number of apps on app downloads.

market with no categories at all. In such markets, the benefits of introducing more categories would be lower, or even negative. This suggests there is likely an optimal number (or range) of categories to display in the mobile app market (conditional on the number of apps), but the variation in my data does not allow me to say what this number (or range) may be. I leave this question to future research.

More generally, in the sense of informing future policy, these results likely represent an upper bound of the magnitude of congestion externalities in online markets. Platforms such as Amazon and Netflix are continuously improving personalization and recommendation algorithms, so that consumers do not need to engage in costly category-based browsing. In a market with better recommendation technology, or a better search function than in the Google Play Store in the 2012-2014 period, the effects of re-categorization, and variety-based congestion, likely would be smaller. That said, despite large investments in targeting and recommendation algorithms, they are far from perfect. For example, despite video streaming services such as Netflix being at the forefront of personalized recommendation technology, the 2019 Nielsen Total Audience Report finds that US consumers spend between 8 and 10 minutes in choosing what to watch. Only a minority of consumers, fewer than one third, watch algorithmically recommended titles ([Nielsen.com](https://www.nielsen.com)). This is likely because consumers have idiosyncratic product preferences, which are difficult for platforms to predict.³⁰ Category browsing is still an important method of consumer discovery, meaning that categorization and variety-based congestion still plays an important role today, and moving forward.

5 Model Based Evidence

In this section, I set up and estimate a demand model that allows me to measure changes in consumer welfare in the market while accounting for congestion costs. I then evaluate the effects of congestion on consumer welfare from additional variety using the re-categorization as an illustrative example.³¹

My demand model conceptually follows the consideration-set framework ([Goeree 2008](#), [Moraga-González et al. 2015](#), [Honka et al. 2017](#)). This framework models the

³⁰It is possible to make an analogy to personalized pricing / first degree price discrimination. Although in principle fully personalized pricing is possible, it is not attempted by firms because of noisy consumer preferences ([Dubé and Misra 2022](#)).

³¹I do this without a formal supply model by relying on reduced form evidence for counterfactual entry patterns. In previous versions of this paper, I introduced an incomplete information static entry model to compute counterfactuals and perform welfare analysis for a sub-sample of free apps. This produced qualitatively similar effects, but required numerous restrictive assumptions.

consumer search process in a setting where consumers are not fully informed about the attributes of each product in the market. Consumers choose a subset of products to “consider,” and then choose a product to consume out of that subset. Using this approach, it is possible to separate two types of product characteristics: (1) characteristics that affect consumer awareness of a product, and the probability a product appears in a consumer’s consideration set (e.g., the difficulty of finding a product; (2) characteristics that affect consumer consumption utility and choice conditional on awareness. I set up and estimate a simple linearized version of such a model below.³² This demand model accounts for fundamental sources of product discovery frictions in the Google Play Store, and in many other online markets. It is consistent with a broad set of implications from theoretical and empirical search literature, and can be credibly estimated with app-level data.

5.1 Demand Model

Consumer i chooses to download app $j \in \{1, 2, \dots, N\}$ in market/period t . The utility she receives from downloading an app j of type c is:³³

$$\begin{aligned} u_{ijct} &= \delta_{jct} + \gamma \ln(N_{c^*(j)t}) + R_{jct}\kappa + \zeta_{ict} + (1 - \sigma)\epsilon_{ijct} \\ &= X_{jct}\beta + \xi_{jct} + \gamma \ln(N_{c^*(j)t}) + R_{jct}\kappa + \zeta_{ict} + (1 - \sigma)\epsilon_{ijct} \end{aligned} \quad (3)$$

where X_{jct} and ξ_{jct} are unobservable and unobservable app/app-type characteristics that affect consumption utility, respectively. Observable app characteristics include its average rating in period t , whether it’s paid or free, app price for paid apps, and app “age” (months on the market). Characteristics that do not enter into δ_{jct} capture utility changes coming from changes in app discovery costs, rather than consumption utility. This distinction is conceptual in the main text, but in the model in Online Appendix D.2, these variables enter the model non-linearly and are distinct from consumption utility.

$\ln(N_{c^*(j)t})$ is the number of apps in the *category* of app j . As discussed in Section 4, this variable captures the congestion externalities of increasing the number of apps in a category on the consumption of app j , by reducing the probability that app j appears in a space that is visible to consumers.³⁴ c^* represents the actual

³²A non-linear version of a consideration-set model following [Moraga-González et al. \(2015\)](#) is in Appendix D.

³³An app’s “type” is defined according to the classification described in Section 3.2.

³⁴This is similar to the modelling in [Akerberg and Rysman \(2005\)](#). $N_{c^*(j)t}$ could also affect

observed categorization of apps in the store rather than app-type (c), although after re-categorization c and c^* coincide for app j .

R_{jct} are additional shifters that proxy discovery costs, capturing competition for unobservable store space and scarce consumer attention. One such shifter is an app's downloads in the previous period - capturing inter-temporal externalities in the mobile app market, and reflecting the design of many online markets. An app with more downloads in period $t - 1$ is more likely to appear on a best-seller list and remain highly visible in period t . R_{jct} also includes time-varying app-type specific fixed effects. Each app-type has two dummies, one before and one after re-categorization.³⁵ Differences in pre/post app-type fixed effects capture average changes in consumer utility from downloading a type c app after re-categorization as compared to before, after conditioning on other observable app characteristics.

ϵ_{ijct} is a consumer/app/market specific demand shock with an iid EVT1 distribution (mean zero, standard deviation normalized to 1).³⁶ ζ_{ict} is a consumer/app-type/market specific demand shock, such that $[\zeta_{ict} + (1 - \sigma)\epsilon_{ijct}]$ is also EVT1 distributed. The consumer/app-type specific shock allows for correlation in consumer preferences across apps within app-types, parametrized by σ . The model otherwise abstracts from unobservable consumer heterogeneity.³⁷ The consumer can also pick an outside option of not downloading anything and receive $u_{i0} = \epsilon_{i0}$.

Based on the assumed distribution of idiosyncratic demand shocks, the market share of app j , belonging to app-type c in market t is defined as:

consumer demand through changes in the number of substitutes rather than through congestion externalities, but reduced form results suggest otherwise. In Section 4, I show that the number of apps in a category matters for downloads when the number of substitutes is held constant. In Online Appendix C.8, I also show that changes in the number of apps in a category do *not* affect paid app prices. If the number of apps in a category primarily changes competition intensity, it should affect paid app prices.

³⁵Formally, $\sum_c \Theta_{c,Pre} (D_c \times D_{Pre}) + \sum_c \Theta_{c,Post} (D_c \times D_{Post})$, where D_c is an app-type c dummy, D_{Pre} is a dummy equal to 1 before March 2014 and D_{Post} is a dummy equal to 1 after March 2014. Θ s are coefficients on the combinations of dummies.

³⁶Consumer preference for variety comes primarily through this shock. This is a reasonable assumption for a market with many minimally horizontally differentiated product versions such as “Angry Birds Space” and “Angry Birds Star Wars.”

³⁷There could be additional unobservable heterogeneity in preferences for quality. Estimating a random coefficients model using aggregate app data with hundreds of thousands of products and a small number of markets is computationally challenging. There is also likely not enough variation in market shares to identify the distribution of unobservable heterogeneity (Berry et al. 2004, Armstrong 2016).

$$s_{jct} = \frac{\exp(\frac{R_{jct}\kappa + \gamma \ln(N_{c^*}(j)_t) + \delta_{jct}}{1-\sigma})}{\sum_{j' \in c} \exp(\frac{R_{j'ct}\kappa + \gamma \ln(N_{c^*}(j')_t) + \delta_{j'ct}}{1-\sigma})} \frac{\left[\sum_{j' \in c} \exp(\frac{R_{j'ct}\kappa + \gamma \ln(N_{c^*}(j')_t) + \delta_{j'ct}}{1-\sigma}) \right]^{1-\sigma}}{\sum_{c' \in \{1, \dots, C\}} \left[\sum_{j'' \in c'} \exp(\frac{R_{j''c't}\kappa + \gamma \ln(N_{c^*}(j'')_t) + \delta_{j''c't}}{1-\sigma}) \right]^{1-\sigma}} \quad (4)$$

where j' is an app that belongs to app-type c and j'' is an app that belongs to app type c' .

5.2 Demand Estimation and Results

Inverting the market share of app j of app-type c at time t produces the following linear estimating equation:

$$\ln\left(\frac{s_{jct}}{s_{0t}}\right) = X_{jct}\beta + \gamma \ln(N_{c^*}(j)_t) + R_{jct}\kappa + \sigma \ln(s_{j|c,t}) + \xi_{jct} \quad (5)$$

where s_{0t} is the market share of the outside option.³⁸ $s_{j|c,t}$ is the within-app-type market share in period t . I estimate the model by solving this equation for the structural error term ξ_{jct} and interacting it with instruments to form the GMM criterion function. As discussed above, X_{jct} includes app-age fixed effects, R_{jct} includes time-varying app-type fixed effects, and I also include year and month fixed effects and developer fixed effects to control for additional unobservable heterogeneity.

The key parameters I look to identify in this model are: the price coefficient for paid apps (one of the variables in X_{jct}), β_{price} , γ and σ . β_{price} is important for measuring the welfare effects of changes to product variety or characteristics in dollar terms. σ identifies consumer substitution patterns across apps, and γ identifies the congestion externality.

The three variables associated with the coefficient of interest suffer from endogeneity. App prices and within-type market shares are likely correlated with unobservable product quality (ξ). Products with higher ξ have more demand, higher prices and higher within-type market shares, meaning that simple OLS estimates of the price coefficient and σ should be biased. Similarly, there may be correlation between the number of apps in a category and some unobservable category/time varying demand shocks, biasing OLS estimates of γ . I address these concerns using instrumental variables.

³⁸I assume that total market size is twice the maximum total number of purchases observed in a time period. The Android handset market is growing over time, so another possible assumption is to match total market size to the number of US Android handsets. Since I include year/month fixed effects, other normalizations do not change parameter estimates qualitatively or quantitatively.

I use characteristics-based instruments to address the endogeneity of app prices and within-group market shares: average ratings, app size (in MB), and number of screenshots of other type c apps, excluding app j . These characteristics proxy for competitors' average quality. For example, the quality of competing apps is excluded from the utility consumers receive from downloading a paid app j , but it does affect app j 's price through competition.

The instrument for $N_{c^*(j)t}$ has to capture supply-side entry cost shocks and exclude demand-side factors. The number of apps in a given category is correlated over time - apps rarely exit, and those who entered in previous periods are persistently present in the market. However, demand shocks could also be correlated over time - consumers may persistently like certain games or game categories in both periods t and $t + 1$. To create my instrument, I estimate a regression of the number of apps in category c^* in period t on the number of apps in category c^* in period $t - 1$, controlling for total category downloads in period t and mean category rating in period t .³⁹ I then use the residual of that regression as the IV. Intuitively, realized period t total downloads and category ratings control for any demand side-shocks that influence entry decisions. The remaining variation in the number of apps between period $t - 1$ and period t should be driven by supply side cost shocks.⁴⁰

I estimate the model using data on all game apps from March 2012 to December 2014 (excluding March 2014). Results for the main parameter of interest are in Table 4.⁴¹ There are four columns in the table. Each column has a different set of instrumental variables. Column (1) estimates an OLS version of the model with no instruments. Column (2) includes characteristics-based IVs for prices and the within-group market share, but no IVs for the number of apps in a category. Column (3) is the opposite: it includes the instrument for the number of apps in a category, but no IVs for prices and within-group shares. Column (4) includes all IVs.⁴² Standard errors are clustered at the app level throughout.

Comparing estimates across columns suggests that the instruments generate appropriate variation to target the relevant endogenous variables. In Columns (1) and (3), σ is very close to 1, which is unintuitive and generally suggests the presence of attenuation bias. The price coefficient is similarly very small, implying that con-

³⁹Estimates of this supporting regression are in Table D2 in Online Appendix D.

⁴⁰This instrument is conceptually similar to the cross-sectional Hausman-Nevo IV specification of Dubois and Lasio (2018).

⁴¹Additional estimates of other parameters that are less important for the counterfactuals in this paper are presented in Appendix D.1.

⁴²In addition to the instruments described in the main text for the main variables of interest, I also instrument for lagged app downloads (a variable in R_{jct}) using functions of earlier lags in app downloads (e.g., q_{jt-2}, q_{jt-3}).

Table 4: Demand Model Parameter Estimates

	(1)	(2)	(3)	(4)
Parameter Estimates				
γ	0.053*** (0.001)	-0.066*** (0.008)	-0.432*** (0.003)	-0.393*** (0.026)
σ	0.963*** (0.000)	0.742*** (0.013)	0.989*** (0.000)	0.709*** (0.027)
β_{price}	-0.001*** (0.000)	-0.355*** (0.051)	-0.001*** (0.000)	-0.836*** (0.111)
Instruments				
Characteristics IVs		•		•
N Apps IV			•	•
Fixed Effects				
Paid App Dummy	•	•	•	•
App Age	•	•	•	•
App Rating	•	•	•	•
Year/Month	•	•	•	•
Developer	•	•	•	•
App Type \times Pre/Post Re-Categorization	•	•	•	•
Other Controls				
Lagged Downloads	•	•	•	•
Other App Quality Proxies	•	•	•	•
Observations	4,584,281	4,190,276	4,152,328	4,152,147

Notes: Sample includes game app-month observations from March 2012 to December 2014 (excluding March 2014) in the Google Play Store. “App Rating FE” are a set of dummies representing the average rating of app j in period t within 0.5 stars. Apps with 2 stars or less are the baseline group for the “App Rating FE.” Year/Month FE include separate dummies for each year (2013 and 2014, relative to a 2012 baseline), and each month (Feb/Mar..., relative to a Jan baseline) in the sample period. Lagged downloads are the downloads of app j in period $t - 1$. Other app quality proxies include the number of screenshots for app j in period t , the size of the app in MB, and a dummy for whether the app has a video preview. Instruments in Columns (2) and (4) include the average ratings of other apps in the same app-type, the average number of screenshots of other apps in the same app-type, and the average size of other apps in the same app-type. Instruments in Columns (3) and (4) include the residual of a regression of the current number of apps in a category on the lagged number of apps in a category (see Table D2 for estimates). Additional IVs in all four columns, for lagged app downloads, include functions of further lags in app j downloads (2 and 3 periods before period t). Standard errors are clustered at the app level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

sumers are price-inelastic and that the vast majority of paid-app marginal costs are negative. This is also problematic, and together, these estimates suggest endogeneity is a substantial concern.

β_{price} and σ estimates in Columns (2) and (4) suggest that characteristics based-IVs appropriately resolve these concerns. In Column (4), the price coefficient implies

a median paid game demand price elasticity of over 2, which is realistically elastic. Estimates of σ in Columns (2) and (4) are 30% smaller than in Columns (1) and (3).

Similarly, the Column (1) estimate for the γ coefficient is positive, which is un-intuitive given the findings in Section 4, and is likely caused by endogeneity. The estimate in Column (2) is negative, but not nearly as negative as in Columns (3) and (4), where the appropriate instrument is incorporated into the estimation procedure. Results in Columns (3) and (4) are consistent with reduced form evidence and suggests that an increase in the number of apps in app j 's category reduces demand and consumer utility. Put otherwise, there are congestion externalities on consumer demand in the mobile app market.

In the welfare analysis below, I use estimates from Column (4), but I also tested several alternative specifications of the model for robustness. Results are in Table D3 in Online Appendix D. For example, I estimate a simple nested logit model with no discovery frictions. This model does not fit the data well, producing σ estimates above 1. I also allow new apps to have different discovery frictions than incumbent apps. There are no statistically significant differences in key parameter estimates between the two types of apps. Last, I estimate a non-linear GMM model based on the search and demand model developed in Appendix D.2. Results are qualitatively similar to those in the main text with respect to the main parameters of interest.

5.3 Welfare Costs of Congestion

In this section, I first show that re-categorization increases entry rates for mobile games and product variety in the market. Then, using the estimated model from the previous section, I calculate the welfare gains consumers experience from the new products and the costs of congestion.

5.3.1 Re-Categorization and Entry

I first test for whether re-categorization affected app entry. In each month, I observe the number of unique entrants of each app-type.⁴³ Similar to Section 4, I estimate the following regression for app type c at time t :

$$\ln(\text{N Entrants}_{ct}) = \tau(\text{Game}_c \times \text{Post}_t) + \delta_c + \delta_t + \epsilon_{ct} \quad (6)$$

⁴³I only consider new apps that appear in the store, rather than apps that switch categories or produce new versions. In general, very few apps switch categories. Less than 0.2% of apps switch categories. A fraction of that percentage switch between being classified as games and non-games.

where δ_c and δ_t are app-type and time fixed effects, $Post_t$ is a dummy equal to one after re-categorization from March 2014, and $Game_c$ is a dummy variable equal to one for the game types and zero for non-game types. The coefficient of interest in Equation 6 is τ , which captures the treatment effect of re-categorization on the outcome. Entry is a long-run decision, so I estimate the regression in Equation 6 using the entire January 2012-December 2014 time period. Since I am looking for aggregate effects, regressions are at the aggregate game/non-game or app-type level.

Table 5 shows results for the two main supply-side outcomes. Columns (1) and (2) use monthly entry by new apps as an outcome variable.⁴⁴

Table 5: Entry Difference in Differences Estimates: Average Effects

<i>Outcome Variable:</i>	ln(N Entrants) (1)	ln(N Type Entrants) (2)
Games \times Post	0.342*** (0.068)	0.556*** (0.095)
Games	-6.914*** (1.893)	
Unit of Observation	Game/Non-Game	App-Type
Time Period	Jan 12 / Dec 14	Jan 12 / Dec 14
Sample	All	All
Year/Month FE	•	•
App-Type FE		•
Observations	70	1,470
R-squared	0.997	0.975

Notes: The sample period in all columns covers January 2012 to December 2014. Sample column (1) includes monthly observations at the Game/Non-Game level. Sample in column (2) includes monthly observations at the app-type level. Column (1) outcome is the natural logarithm of number of new games/non-games on Android. Column (2) outcome is the natural logarithm of the number of new apps in a given game/non-game *app-type* on Android. Controls include year and month fixed effects, game/non-game fixed effects for odd columns, and app-type fixed effects for even columns. Additional controls include app-type specific time trends. The variable “Games \times Post” is a dummy variable equal to 1 for games (or game app-types for even columns) during and after March 2014. Standard errors are robust to heteroskedasticity in odd columns and clustered at the app-type level in even columns. *** p<0.01, ** p<0.05, * p<0.1

Entry treatment coefficients are 0.34 at the game/non-game level and 0.56 at the app-type level. The estimates are statistically significant at the 99% confidence level. They show that following the re-categorization, developers entered 34% more game apps than non-game apps. At the app-type level, entry increased by over 50% for game app-types relative to non-game app-types.⁴⁵

⁴⁴This measure only counts apps that have previously not existed in the store and became active. Existing apps that changed categories or that had updates are not counted in this measure.

⁴⁵See Table C9 in the Online Appendix for absolute entry results, which estimate that an addi-

In Appendix Figure C3, I show that the effects begin two months after the re-categorization announcement in December 2013. This is consistent with the average timing of app development, and suggests that new apps enter in response to the re-categorization. Figure C3 also shows that entry changes peak the month after re-categorization, but are present and persistent throughout the post-policy period. Appendix C.7.1 shows the mechanism through which the entry effects operate. Entry is primarily driven by the new categories in the store, which improved consumers' information and reduced discovery costs. Entry is also bigger for app-types which saw greater immediate reductions in congestion.⁴⁶

Over all, the estimated entry changes reflect theoretical predictions from search models, such as Anderson and Renault (1999) and Chen and Zhang (2017). As product discovery costs fall, consumers are more likely to search more, and producers respond by entering new products to benefit from increased visibility. Results are also consistent with previous descriptive comparisons between high and low discovery cost markets. Brynjolfsson et al. (2003) shows that online retailers have between 10 and 30 times more products than brick-and-mortar retailers. Aguiar and Waldfogel (2018) notes that the number of new music products between 2000 and 2008 tripled. However, neither of these papers separates changes in search costs from other changes in technology, such as production costs.

The increase in entry should satisfy consumer preferences for variety, but it should also increase congestion externalities. Based on demand parameter estimates in Section 5.2, this should restrict the benefits consumers receive from the new products. I quantify these welfare effects below.

5.3.2 Welfare Decomposition

Based on the demand model in Section 5.1, changes in consumer surplus are the difference between expected utilities under different sets of products. I convert these changes into dollar values using the price coefficient from Table 4. A consumer who downloads app j of type c receives baseline utility $\hat{\delta}_j = X_j \hat{\beta}$. That consumer also receives/pays congestion costs based on the number of other apps in the category ($\hat{\gamma} \ln(N_{c^*(j)t})$) and other "discovery costs" $R_{jct} \hat{\kappa}$. The last two terms are linearly additive to $\hat{\delta}_j$. Based on the demand model from Section 5.1, expected consumer

tional 1,500 game apps enter the average game app-type after re-categorization.

⁴⁶As for the downloads results in Section 4, these estimates are particular to the small baseline number of app categories. As discussed in Section 4.2, in a market with many categories, such as Amazon, more consumer time and effort would be spent browsing *between* categories, as compared to *within* categories. There, the effect of introducing more categories on entry would likely be smaller.

utility is:

$$EU = \ln \left(1 + \sum_{c \in \{1, \dots, C\}} \left[\sum_{j \in \Omega_c} \exp \left(\frac{\hat{\delta}_j + R_{jc} \hat{\kappa} + \hat{\gamma} \ln(N_{c^*(j)})}{1 - \sigma} \right) \right]^{1-\sigma} \right) \quad (7)$$

where Ω_c is the set of apps of type c and $\Omega = \{\Omega_1, \Omega_2, \dots, \Omega_c, \dots, \Omega_C\}$.

I calculate three relevant expected utility measures. First, I calculate expected utility using *the actual set of apps available in December 2014*, the last period available in the sample. I refer to this expected utility as EU_1 . Second, I calculate expected utility using a set of apps in a *counterfactual December 2014 market without re-categorization*, EU_2 . I calculate this counterfactual choice set by randomly removing apps that entered between March 2014 and December 2014 from the December 2014 choice set.⁴⁷ Reduced form estimates from Section 4 tell me how many entrants to remove. I remove entrants to match the aggregate entry treatment effects from Column (1) of Table 5.⁴⁸ I repeat this exercise 500 times to ensure randomization is not driving results. It is important to note that for a given product j , there is no difference in $\hat{\delta}_j$ or in R_{jc} between the factual and counterfactual markets. Only the set (and number) of products in the market changes.

The difference $\frac{EU_1 - EU_2}{-\beta_{price}}$ is the overall effect of additional entry on consumer welfare. This is a “net” effect. Re-categorization changes Ω , the set of products in the market, and also $N_{c^*(j)}$, the congestion measure. To separate the two, I calculate EU_3 , which is based on Equation 7 where Ω comes from the factual December 2014 market, and $N_{c^*(j)}$ comes from the counterfactual December 2014 market. Put otherwise, it measures expected consumer utility under the factual choice set, but with counterfactual congestion based on the counterfactual number of entrants. $\frac{EU_3 - EU_2}{-\beta_{price}}$ isolates the *gross* entry effect of only changing the choice set while holding congestion constant. $\frac{EU_1 - EU_3}{-\beta_{price}}$ isolates the negative welfare effect due to the increase in congestion between the counterfactual and factual markets.

⁴⁷An alternative approach to doing this involves setting up and estimating a supply model, such as a dynamic incomplete information app entry model. With estimates from this model, it is possible to compute equilibrium demand and supply under different store categorizations. I do not take this approach as it would require me to impose numerous strong assumption. For example, I would have to introduce assumptions about the information structure of entrants in the market, fixed and entry costs, and product design and location decisions by entrants. Estimating an entry model with many players also introduces severe computational challenges.

⁴⁸In aggregate, 146,688 game apps entered between March 2014 and December 2014. Based on Column (1) of Table 5, factual entry is 34% higher than without re-categorization. This means that $\frac{146,688}{1+0.34} = 109,469$ game apps would have entered the market without re-categorization, and I need to remove $146,688 - 109,469 = 37,219$ entrants.

Table 6 shows the decomposition of consumer surplus changes. I show mean entry and congestion effects based on the randomized procedure outlined above.⁴⁹

Table 6: **Estimated Changes to Monthly Per-Consumer Surplus (\$)**

1) Mean Gross Entry Effect	$EU_3 - EU_2$	0.057
2) Mean Entry Congestion Effect	$EU_1 - EU_3$	-0.024
3) Net Entry Effect	$EU_1 - EU_2$	0.033

Notes: This table shows estimates of average per-consumer welfare effects using demand model estimates from Table 4. Numbers shown are averages from 500 simulations. The full distributions of simulation effects are in Figure D3 in the Online Appendix.

Without accounting for congestion externalities, the additional entry spurred by the re-design of the store has a large effect on consumer welfare. Row (1) of Table 6 shows that *gross* consumer welfare increases by 5.7 cents per consumer per month, adding up to 68.4 cents of additional per-consumer welfare in a year. This is equivalent to approximately one third of the price of the average paid game app. Across all 100 million Google Play Store consumers in 2014, gross annual welfare gains add up to nearly \$70 million.

However, row (2) of Table 6 shows that the congestion externalities associated with the increase in product variety reduce consumer welfare gains. Consumer lose approximately 40% of gross variety welfare gains because of increasing variety-driven congestion costs. In absolute terms, losses from additional congestion are 2.4 cents per-consumer per-month relative to a counterfactual world where re-categorization does not happen and entry does not increase. The costs add up to approximately \$29 million per-year for all Google Play Store users.

Over all, consumer welfare still increases. Row (3) shows that the net increase in consumer surplus is 3.3 cents per-consumer per-month, or approximately 40 cents per-year (and \$40 million across all consumers), relative to a counterfactual world where re-categorization does not happen. Nonetheless, these estimates show that congestion externality costs in online markets are economically substantial, and consumers do not fully benefit from additional product variety in the market.

⁴⁹Figure D3 in the Online Appendix shows the full distributions of these effects under different random draws. The variation in effect size across different randomization seeds is quantitatively small.

5.4 Model and Counterfactuals Discussion

The demand model and counterfactuals presented above (and in Appendix D) abstract from several elements of the consumer search process in the Google Play app store. It does not explicitly model the costs consumers incur while browsing *between* categories, only the costs of browsing *within* categories. In a market with many categories, consumers may spend a substantial amount of time choosing which category to look through. I lack consumer level data to be able to estimate such costs, but it is reasonable to assume that in the mobile app market in 2012-2014, they are likely very small compared to the costs of browsing within categories. They also do not change for the relevant counterfactual welfare decomposition in Section 5.3. Some estimates from the model speak to their magnitude/ importance: the model measures average changes in consumer utility from downloading a type c app before and after the re-design using time-varying app-type fixed effects. Within app-type differences in these fixed effects capture many changes, including better consumer information about category content, but they should also include changes in category browsing time/ costs. If the costs of browsing between categories become large and important after the re-design, the difference in fixed effects should be negative. However, estimates of these differences in Appendix Figure D1 are uniformly positive. This suggests that category browsing costs are either small in magnitude, or that they do not increase substantially. In that sense, abstracting from these in the model should not affect the main results and conclusions for this particular setting.

The model also abstracts from platform personalization on the Google Play Store. In reality, all consumers see personalized app recommendations on the Google Play Store home page. Some consumers may immediately see a targeted product that caters to their preferences, and purchase it immediately without engaging in a costly browsing process. Other consumers may find what they are looking for quickly using the search bar function. For those users, variety-driven congestion does not matter. However, based on consumer surveys (see Appendix A.3), and on statements by industry experts (see Section 2), this was likely a minority of app consumers during (and even after) my sample period. Abstracting from this element should not substantially affect my estimates or conclusions for this market.

These issues also raise broader questions about the extent to which the findings here generalize. There are two possible generalizations to be made from the findings of this paper. First, my results suggest that past findings of the value of added product variety online (e.g., Brynjolfsson et al. 2003 and Aguiar and Waldfogel 2018) may be overstated, because they do not account for congestion externalities. Papers in this literature primarily use pre- 2012 data from online markets, where targeting and personalization technology was plausibly worse than in the Google Play Store.

The magnitude of my estimated congestion externalities, therefore, is appropriate, and could even be a lower bound.

Second, my findings have implications for future platform- led or regulator- led policies that affect online entry. These could involve platform re-design and re-categorization, as in my illustrative case, or other policies such as direct subsidies to entrants.⁵⁰ A policy increasing entry would increase consumer welfare through new variety, but the benefits would be mitigated by increased congestion. Future policies are likely to occur on platforms with substantially better search and personalization technology than the Google Play app store in 2012-2014. As such, the counterfactual results suggesting that congestion costs dissipate 40% of welfare gains from additional variety are likely an upper bound. At the same time, personalization technology is unlikely to ever be perfect, as consumers have idiosyncratic product preferences that are difficult to predict.⁵¹ As discussed in Section 4.2, recent surveys show that streaming video consumers still spend substantial time browsing platforms, and that most consumers do not choose the algorithmically provided option. Since streaming video platforms such as Netflix are at the forefront of personalization and recommendation technology, this implies that variety-driven congestion should still have an effect today and in the future.

6 Conclusion

This paper examines the role of variety-based congestion in online markets. Product assortment is the key competitive outcome of concern for regulators of online markets. Understanding consumer welfare changes in response to changes in assortment is important. I use new data from the Google Play mobile app store, taking advantage of a re-design of the game section of the store. Reduced form estimates using changes in downloads show that when the re-design increased the number of categories, reducing the number of apps-per-category, downloads increased. Downloads also increased by more for apps that were in less populous categories. This points to the existence of variety-based congestion externalities.

I set up a structural demand model that accounts for variety-driven congestion externalities. This allows me to decompose consumer welfare changes from variety into “gross” changes in utility due to the larger assortment, and welfare changes driven by congestion. Increased mobile game entry following the Google Play Store

⁵⁰Alibaba, for example, provided direct subsidies and free services to firms operating on its platform (ChinaDaily.com.cn).

⁵¹As discussed previously, this is analogous to the inability of firms to first-degree price discriminate online ([Dubé and Misra 2022](#)).

re-design serves as an illustrative example. Entry increased by 34% after the re-design, because of improvements in consumer discovery technology. Relative to a counterfactual market where the re-design did not happen, each consumer received an additional 40 cents per-year from the new product assortment. This adds up to nearly \$40 million additional aggregate welfare per year. However, absent congestion, welfare increases would have been even higher. Increased congestion takes away 40% of gross variety gains. These findings suggest that previous estimates in the literature were overstating consumer benefits from additional online product variety.

Policy-makers are concerned with platforms foreclosing potential entrants and reducing variety, for example, by making product discovery difficult for consumers ([WSJ.com](https://www.wsj.com)). My findings show that welfare gains from variety are counteracted by increasing congestion externalities. In my application, welfare changes coming from congestion are not enough to overwhelm the variety effect, and consumers benefit. However, welfare changes from increased congestion are still economically significant and annually add up to tens of millions of dollars of lost consumer surplus. Personalization and recommendation technology has improved since my sample period, and it is substantially better in many online markets today (e.g., Amazon, video streaming services). Nonetheless, these findings serve as a useful upper bound on the magnitude of congestion effects that could be generated by future platform-led or regulator-led policy aiming to increase entry online. Empirical analysis of platform incentives to invest in better search and personalization technology - or change their design to facilitate consumer discovery - in the presence of congestion externalities, should be a fruitful direction for future research.

References

- Akerberg, D. A. and M. Rysman (2005). Unobserved product differentiation in discrete-choice models: estimating price elasticities and welfare effects. *RAND Journal of Economics* 36(4), 771–789.
- Aguiar, L. and J. Waldfogel (2018). Quality predictability and the welfare benefits from new products: Evidence from the digitization of recorded music. *Journal of Political Economy* 126(2), 492–524.
- Anderson, S. P. and R. Renault (1999). Pricing, product diversity, and search costs: a bertrand-chamberlin-diamond model. *RAND Journal of Economics* 30(4), 719–735.

- Armstrong, T. B. (2016). Large market asymptotics for differentiated product demand estimators with economic models of supply. *Econometrica* 84(5), 1961–1980.
- Bar-Isaac, H., G. Caruana, and V. Cuñat Martinez (2012). Search, design and market structure. *American Economic Review* 102(2), 1140–1160.
- Berry, S., O. B. Linton, and A. Pakes (2004). Limit theorems for estimating the parameters of differentiated product demand systems. *The Review of Economic Studies* 71(3), 613–654.
- Brynjolfsson, E., Y. Hu, and D. Simester (2011). Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales. *Management Science* 57(8), 1373–1386.
- Brynjolfsson, E., Y. Hu, and M. D. Smith (2003). Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers. *Management Science* 49(11), 1580–1596.
- Chen, Y. and T. Zhang (2017). Entry and welfare in search markets. *The Economic Journal* 128(608), 55–80.
- Dinerstein, M., L. Einav, J. Levin, and N. Sundaresan (2018). Consumer price search and platform design in internet commerce. *American Economic Review* 108(7), 1820–59.
- Dubé, J.-P. and S. Misra (2022). Personalized pricing and customer welfare. *Journal of Political Economy*.
- Dubois, P. and L. Lasio (2018). Identifying industry margins with price constraints: Structural estimation on pharmaceuticals. *American Economic Review* 108(12), 3685–3724.
- Fradkin, A. (2017). Search, matching, and the role of digital marketplace design in enabling trade: Evidence from airbnb. *Matching, and the Role of Digital Marketplace Design in Enabling Trade: Evidence from Airbnb (March 21, 2017)*.
- Gandhi, A., Z. Lu, and X. Shi (2014). Demand estimation with scanner data: Revisiting the loss-leader hypothesis. mimeo.
- Goldmanis, M., A. Hortaçsu, C. Syverson, and Ö. Emre (2010). E-commerce and the market structure of retail industries*. *The Economic Journal* 120(545), 651–682.

- Kesler, R., M. Kummer, and P. Schulte (2020). Competition and privacy in online markets: Evidence from the mobile app industry. In *Academy of Management Proceedings*, Volume 2020, pp. 20978. Academy of Management Briarcliff Manor, NY 10510.
- Liebowitz, S. J. and A. Zentner (2020). The challenges of using ranks to estimate sales. *Available at SSRN 3543827*.
- Liu, Y., D. Nekipelov, and M. Park (2014). Timely versus quality innovation: The case of mobile applications on itunes and google play. *NBER Working Paper*.
- Quan, T. W. and K. R. Williams (2018). Product variety, across-market demand heterogeneity, and the value of online retail. *The RAND Journal of Economics* 49(4), 877–913.
- Zentner, A., M. Smith, and C. Kaya (2013). How video rental patterns change as consumers move online. *Management Science* 59(11), 2622–2634.

APPENDIX TO: “Variety-Based Congestion in Online Markets: Evidence from Mobile Apps” (for Online Publication)

A Appendix A

A.1 List of Non-Game Categories

Table A1: Google Play Non-Game Categories

Books & Reference	Libraries & Demo	Productivity
Business	Lifestyle	Shopping
Comics	Media & Video	Social
Communications	Medical	Sports
Education	Music & Audio	Tools
Entertainment	News & Magazines	Transportation
Finance	Personalization	Travel & Local
Health & Fitness	Photography	Weather

A.2 Examples of Consumer Responses to Re-Categorization

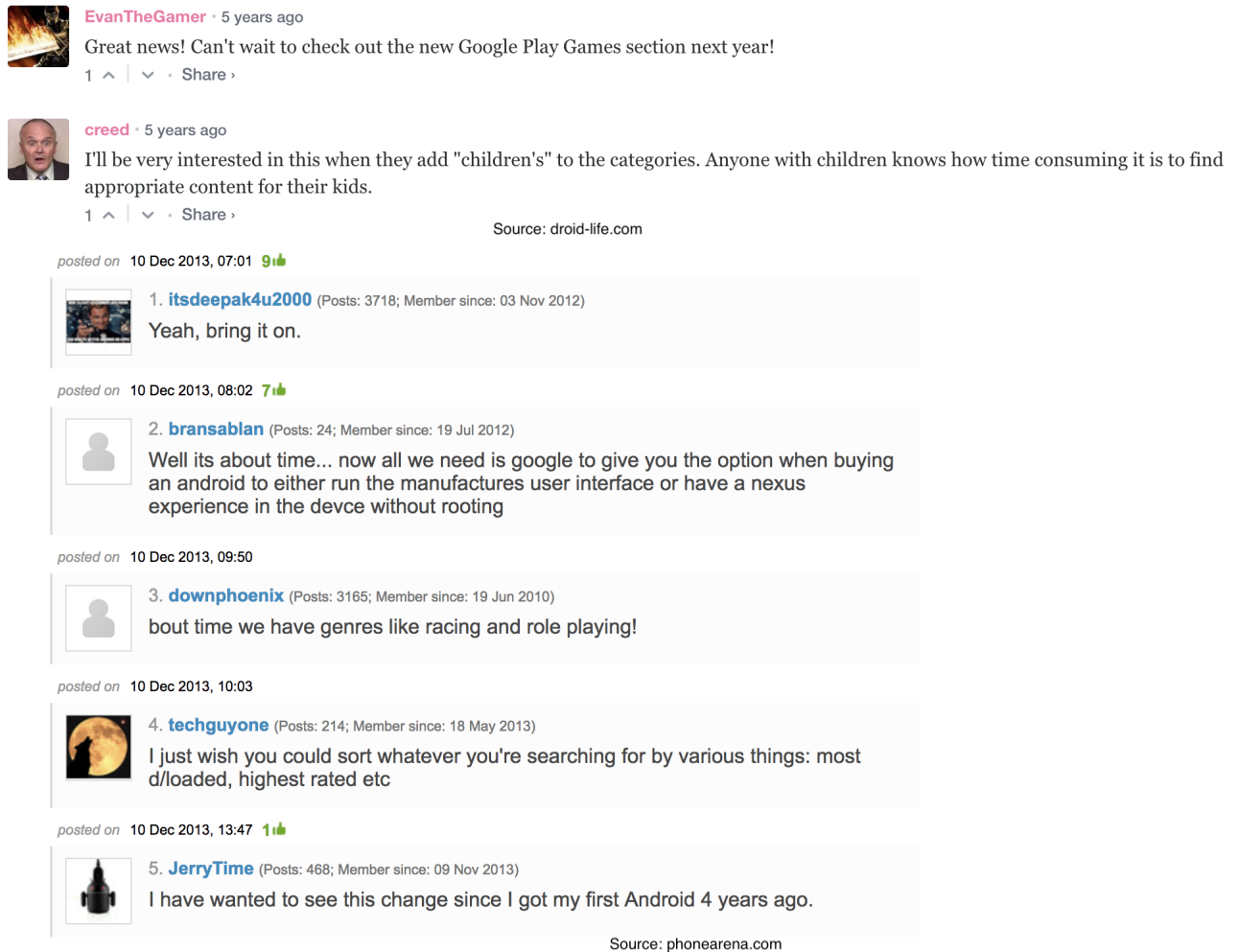


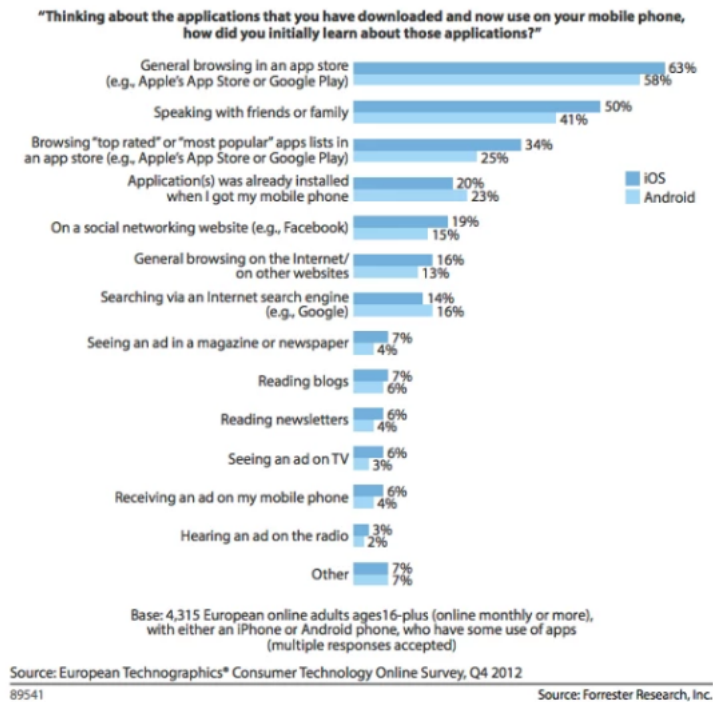
Figure A1

A.3 Consumer App Discovery Surveys

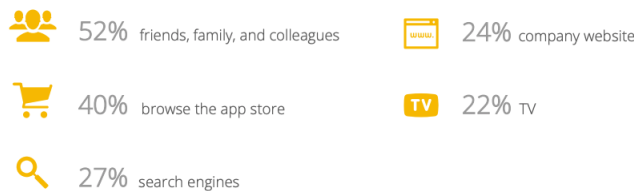
Figure A2 shows results from two consumer surveys taken during my sample period by Forrester and Nielsen. Both surveys asked app consumers how they discovered new products. 58% of Android consumers discover new products through “General browsing in the app store”, according to Forrester, and 25% discover new products

through more targeted browsing - looking at “top rated” or “most popular” app lists in the app store. Only a small share of consumers discover new apps through an internet search engine. Answers are similar in the Nielsen survey, with 40% of consumers browsing the app store to discover new products and only 27% using search engines.

Figure A2: Google Play Consumer Product Discovery Surveys



Sources of awareness of smartphone apps:



Base: Total respondents: Vertical average (n=8,470)
Google/psos Survey Q11. In which of the following ways have you first become aware of [...] smartphone apps? Please include all the sources where you have seen or heard information about apps, even if you didn't subsequently download them.

thinkwithgoogle.com

Figure A3 displays results from two additional app consumer surveys. The top panel shows results from a Nielsen survey from 2011. Over 60% of surveyed con-

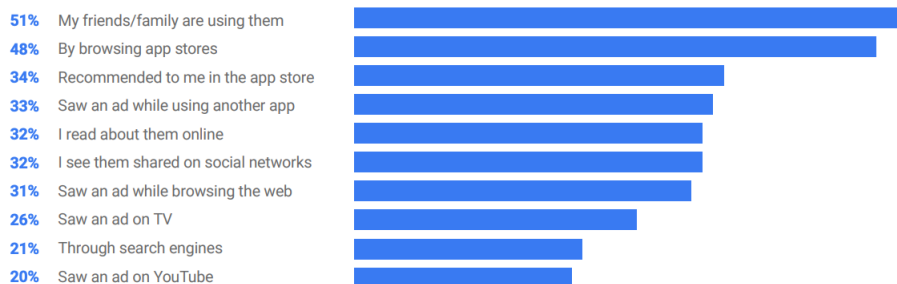
sumers on both Android and the iOS app store discover new products by searching the app store. This is not defined as “browsing” and could include using the search function of the store. It is a far more popular method of discovery than by going through other 3rd party sites or advertising.

The bottom panel of Figure [A3](#) shows results from a 2016 Google and Nielsen survey. Once again, besides learning about apps from friends and family, the most popular discovery method is browsing the app store. This survey is from nearly two years after the end of my sample period and follows substantial improvements in the integration of apps into Google search results. Still, only 21% of users discover new apps using search engines.

Figure A3: Additional Google Play Consumer Product Discovery Surveys



Top methods of app discovery



Base: 999
Q32. How do you typically find out about new smartphone apps?
Q33. And have you found out about an app in any of these ways?

think with Google

B Appendix B

B.1 Data Management - Classifying App Types Using Text

I use a Random Forest machine learning algorithm that first maps the descriptions of the classified post-March 2014 apps into categories, and then projects this mapping on other apps. After removing “stopwords” (e.g., “and”, “or”) I convert app

descriptions into vectors of words and then into term frequency-inverse document frequencies. This method assigns the highest weight to words appearing frequently in an app’s description relative to the average description. I use April 2014 apps as the training set for a Random Forest classifier (other classifiers such as KNN give similar result). I then apply the classifier to apps in every month prior to March 2014. This is similar to how [Liu et al. 2014](#) map Google Play categories into Apple iTunes categories.

B.2 Data Management - Predicting App Downloads

Raw app data includes a range of cumulative downloads that an app accrues over its lifetime. The full list of download ranges is in Table B1. This range is observable in every snapshot of the store. It is conceptually straight-forward to define “per-period downloads” as the difference in lifetime downloads between period t and period $t - 1$. For example, the difference in the lower bounds of lifetime downloads, or in the average of lifetime downloads.

Table B1: List of Cumulative Download Ranges

Lower Bound	Upper Bound
1	5
5	10
10	50
50	100
100	500
500	1,000
1,000	5,000
5,000	10,000
10,000	50,000
50,000	100,000
100,000	500,000
500,000	1 million
1 million	5 million
5 million	10 million
10 million	50 million
50 million	100 million
100 million	500 million
500 million	1 billion

However, the bandwidth increases with the number of downloads, starting at 4 downloads ([1-5], [5-10]) and increasing to 40 ([10-50]) and eventually to 400 million ([100 million - 500 million]). This introduces two possible sources of measurement error, which become worse for more successful apps: (1) overestimation of per period downloads for apps that move from one level to another. An app with a range of [100 thousand - 500 thousand] downloads that moves to the [500 thousand- 1 million] range in the next period could have been downloaded 500 thousand times or 3 times. (2) underestimation of per period downloads for successful apps. An app in the [100 million - 500 million] download range can have millions of downloads every week and stay in the same range.

I rely on two features of the data to recover weekly or monthly app downloads. First, the lifetime download bandwidth for *new entrants* is equal to the per-period bandwidth: an app that entered one period ago and is in the 10 thousand to 50 thousand range was downloaded between 10 thousand and 50 thousand times in the period. Second, I observe weekly category rankings which reflect the 500 most-downloaded apps in each category roughly over the past week.¹ At a weekly frequency, the rankings and downloads of new apps are known. Summary statistics are in Table B2.² I use these apps to predict the downloads of other apps in the market.

Several studies of online markets with best-seller lists find that the Pareto distribution accurately characterizes the rank-downloads relationship (Chevalier and Mayzlin 2006; Garg and Telang 2013).³ The Pareto distribution is a negative exponential distribution where an app at rank n has exponentially more downloads than the app at rank $n + 1$. I fit this distribution for every week and category by estimating an OLS regression of the logarithm of the rank of paid or non-paid ($p \in \{Paid, Non - Paid\}$) new app j in category c at week a of month t on the logarithm of the downloads for every category and week:

¹It is not precisely known how the lists are determined, but Google releases (AdWeek.com) as well as anecdotal industry evidence (Quora) suggest that they reflect the downloads of apps over the previous several days.

²I can assign the lower bound of the bandwidth as the number of weekly downloads, the upper part of the bandwidth, or the average of the bandwidth. In the rest of the analysis of this paper I assign the lower end of the bandwidth, since the average and upper parts of the bandwidth produce unrealistic estimates of downloads.

³It is possible that the Pareto distribution does not correctly predict downloads in this market. Eeckhout (2004) shows that the Pareto distribution accurately predicts the rank-size relationship for the upper tail of the distribution but not necessarily for the lower tail. Liebowitz and Zentner (2020) similarly shows evidence of potential inaccuracy in approximations using distributional assumptions. I use an alternative measure of downloads that does not rely on the Pareto distribution in Online Appendix C.4.3. I also estimate my main results only for new apps, which are not affected by distributional assumptions. Results are qualitatively similar across the two sale proxies.

$$\ln(\text{Downloads}_{jcat}) = \delta_{cp} + \delta_{tp} + \beta_1 \ln(\text{Rank}_{jcat}) + \beta_2 \ln(\text{Rank}_{jcat}) \times \text{Paid}_j + \mu_{jcat}$$

where δ s are category and year/month dummies,⁴ and where μ_{jcat} is a mean zero random variable representing measurement error. β s are slope coefficients that differ for paid and non-paid apps.⁵ I use the lower bound of the bandwidth (minimum downloads in a week) as the dependent variable.⁶ Summary statistics for new apps are in Table B2 and regression estimates are in Table B3. Estimated Pareto Distribution parameters are broadly consistent with similar exercises in the literature (Garg and Telang 2013; Leyden 2018).

I predict the downloads of all apps in the market with estimates from this regression. Only the top 500 ranks each week are observed. To generate rankings for the unranked apps, I sort them based on their number of cumulative lifetime downloads and their age in every week and break up ties by randomizing.⁷

This prediction algorithm depends on variation in app rankings over time. New apps should be able to enter into the rankings at different points in the distribution for me to estimate the Pareto relationship accurately. This is true in the data. While a large proportion of apps not change their rankings from week to week, many apps move at least two spots on a weekly basis. Figure B1 shows the distribution in weekly changes in game rankings.

B.3 Summary Statistics

⁴ δ_{cp} represents $\delta_{cp} \sum_c (D_c \times \text{Paid}_j + D_c \times (1 - \text{Paid}_j))$, where D_c is a dummy for whether j belongs to category c and Paid_j is a dummy for whether app j is paid. δ_{tp} represents $\delta_{tp} \sum_t (D_t \times \text{Paid}_j + D_t \times (1 - \text{Paid}_j))$, where D_t is a dummy for month t .

⁵Predictions do not change substantially when slope coefficients also vary by time or category. I also test for the heterogeneity of the slope coefficients by app-ranking in Column (2) of Table B3, and do not find statistically significant differences in slope between very high ranked and lower ranked apps.

⁶Results using the upper bound or the average clearly overstate the number of downloads. For example, the model predicts each of the top 50 apps to have over 10 million weekly downloads.

⁷To check that randomization does not drive any of the main estimates, I re-estimate the analysis several times with different randomized seeds. The results remain qualitatively and quantitatively similar.

Table B2: Summary Statistics of New Apps at Weekly Level

Variable	Mean	Std. Dev.	Min	Max	N Obs
Games					
Download Lower Bound	19,035	202,837	1	1 million	15,958
Non-Games					
Download Lower Bound	8,473	322,478	1	5 million	28,699

Table B3: Regression Results on Downloads

Outcome Variable:	$\ln(\text{Min Downloads Bound})$		
	(1)	(2)	(3)
	Games	Non-Games	
$\ln(\text{Rank})$	-0.973*** (0.074)	-0.979*** (0.080)	-0.981*** (0.046)
$\ln(\text{Rank}) \times \text{Paid}$	-1.170*** (0.060)	-1.112*** (0.070)	-1.016*** (0.084)
$\ln(\text{Rank}) \times \text{Low-Ranked}$		-0.010 (0.031)	
$\ln(\text{Rank}) \times \text{Paid} \times \text{Low-Ranked}$		0.104 (0.066)	
Year/Month FE	•	•	•
Year/Month FE \times Paid	•	•	•
Category FE	•	•	•
Category FE \times Paid	•	•	•
Observations	15,958	15,958	28,699
R-squared	0.754	0.754	0.802

Notes: The sample in Columns (1) and (2) are new games (games in their first week on the market). The sample in Column (3) are new non-games (non-game apps in their first week on the market). The outcome variable in both columns are the log of the lower bound of the number of weekly downloads for the apps. “Low-Ranked” apps are apps ranked below 50. Controls include year/month and category fixed effects interacted with a paid app dummy. (Category \times Paid) clustered standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

C Appendix C

C.1 Downloads: Average Effects

Figure B1: Weekly Changes in App Ranking on Top 500 Best-Seller Lists

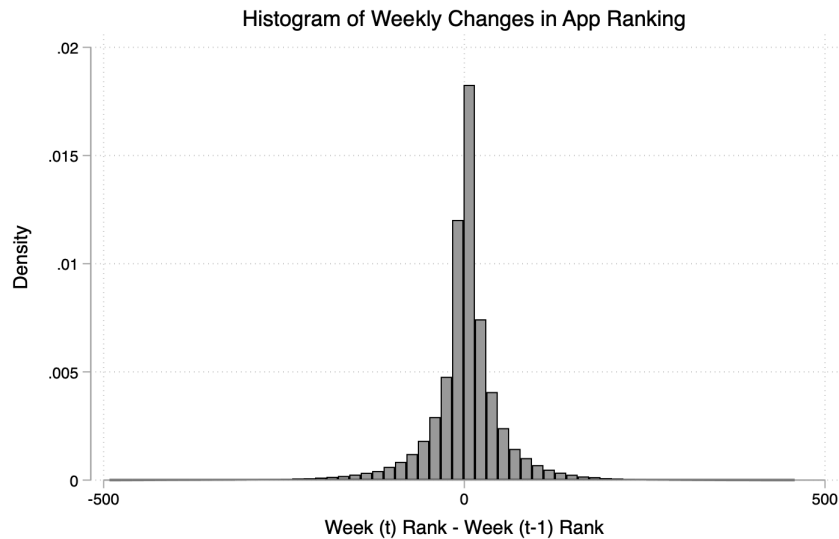


Table B4: Summary Statistics at the App Type-Month Level

Variable	Mean	Std. Dev.	N
Game Types			
Number of Apps	7,552	13,386	630
Number of New Apps	720	1,221	630
Non-Game Types			
Number of Apps	33,764	30,960	840
Number of New Apps	2,710	3,026	840

Table C1: Downloads Difference in Differences Estimates: Average Effects

<i>Outcome Variable:</i>	ln(Tot. Downloads) (1)	ln(Tot. Type Downloads) (2)	ln(App Downloads) (3)	ln(Tot. Type Downloads) (4)	ln(App Downloads) (5)
Games \times Post	0.748*** (0.112)	1.275*** (0.203)	0.162* (0.090)	1.399** (0.288)	0.438* (0.167)
Games	-7.423*** (2.860)				
Unit of Observation:	Agg. Game/Non-Game	App-Type	App	App-Type	App
Time Period:	Jan 12/Dec 14	Jan 12/Dec 14	Jan 12/Dec 14	Jan 14/Apr 14	Jan 14/Apr 14
Sample:	All	All	All	All	All
Year/Month FE	•	•	•	•	•
App-Type FE		•		•	
App FE			•		•
App Controls			•		•
Observations	70	1,470	32,964,682	168	5,284,311
R-squared	0.970	0.929	0.953	0.780	0.972

Notes: The sample period in the first three columns is January 2012-December 2014 and in the last two columns is January 2014-April 2014. All app types and apps are considered in each sample. Data in Column (1) consists of monthly observations at the Game/Non-Game level. Data in Columns (2) and (4) consists of monthly observations at the app-type level. Data in Columns (3) and (5) consists of monthly observations at the app level. Outcomes are natural logarithms of downloads at each aggregation level. Controls include year and month fixed effects, game/non-game fixed effects, app-type fixed effects, or app fixed effects, depending on the column. Additional controls in Columns (1)-(3) include game/non-game or app-type specific time trends. Additional app-level controls for Columns (3) and (5) include average app ratings, a dummy for whether the app is free or paid, the price of the app if it is paid, and app age-specific fixed effects. The variable “Games \times Post” is a dummy variable equal to 1 for games, or game app-types/apps starting from March 2014. Standard errors are robust to heteroskedasticity in Column (1) and are clustered at the app-type level in the remaining columns. *** p<0.01, ** p<0.05, * p<0.1

C.2 Downloads: Category Informativeness Mechanism

After re-categorization, the titles of game categories potentially became more informative about the types of apps present. Before the new categories, consumers looking for music, family or strategy games did not know precisely where to look. After re-categorization, this changed. Consumers had clear information about where different types of apps were located in the store. On the supply side, developers knew that if they produce such a game, there is a clear place for it to be discovered.⁸

I test for this effect in the data. Eight of the eighteen app-types had visibility as game categories before the change: Action and Arcade were grouped together as Arcade & Action, Card and Casino were grouped as Card & Casino. Puzzle was named Brain & Puzzle, and Racing, Sports and Casual games remained unchanged. The remaining app-types did not have pre-existing categories: Adventure, Board, Education, Family, Music, Role Playing, Simulation, Strategy, Trivia and Word games.⁹ Consumers should have become much better informed about the location of these app-types after re-categorization and should be able to reach them much faster. If category informativeness plays a role in discovery frictions, downloads for app-types without pre-existing categories should be more affected by re-categorization than app-types with pre-existing categories. I estimate the following regression at the app-type and app level:

$$y_{(j)ct} = \tau^1 \text{Post}_t \times \text{Game}_c + \tau^2 \text{Post}_t \times \text{Game}_c \times \text{No Pre-Existing}_c + \delta_{(j)c} + \delta_t + e_{(j)ct} \quad (1)$$

This regression is estimated using the four months around the re-categorization event (January-April 2014).¹⁰ δ_t and $\delta_{(j)c}$ are month and app-type or app fixed effects, Post_t is a dummy equal to one in March and April 2014, Game_c is a dummy equal to one for all app-types or apps that are games, and No Pre-Existing_c is a dummy equal to one for the ten app-types that did not have categories before March 2014 and zero otherwise. Non-game app-types/apps are the baseline group.¹¹

⁸This is especially the case since the new categorization structure already existed on the Apple store for years at that point and developers frequently produce apps for both platforms.

⁹Even though they did not exist in the Google Play store, these were categories in the Apple iOS app store for several years prior.

¹⁰I also estimate it using only February and March 2014 in Table C3. Results are qualitatively similar but quantitatively smaller.

¹¹It is possible that there are some informativeness effects even for app-types with pre-existing categories. For example, consumers searching for Casino apps know after re-categorization that there are only Casino-type apps in the Casino category. Apps belonging to these types also experience changes in the number of other apps in their categories and congestion, which also affects downloads as I show below.

Results from these regressions are in Table C2. They show that app-type-level and app-level downloads increase more for games that did not have pre-existing categories. There is a 44 percent increase in downloads after re-categorization for app-types with pre-existing categories, but the change for an app-types without pre-existing categories is four times as large. The heterogeneity is similar at the app level after controlling for app fixed effects. In Columns (3) and (4), I restrict the sample by excluding some game and non-game types that are very different than game types without pre-existing categories.¹² I still find similar heterogeneity in effects. In Table C4 I also show statistically null effects in response to non-existent events taking place before and after actual re-categorization. These results suggest that re-categorization made the category structure more informative and reduced consumer discovery frictions, increasing downloads.

Table C2: Downloads Difference in Differences Estimates: Category Informativeness

<i>Outcome Variable:</i>	ln(Tot. Type Downloads) (1)	ln(App Downloads) (2)	ln(Tot. Type Downloads) (3)	ln(App Downloads) (4)
Games \times Post	0.442* (0.164)	0.220 (0.120)	0.545** (0.134)	0.476 (0.215)
Games \times Post \times No Pre-Existing	1.723*** (0.241)	1.987*** (0.191)	1.635*** (0.267)	1.665*** (0.195)
Unit of Observation:	App-Type	App-Type	App	App
Sample Period:	Jan 14/Apr 14	Jan 14/Apr 14	Jan 14/Apr 14	Jan 14/Apr 14
Sample:	All	All	Small Types	Small Types
Year/Month FE	•	•	•	•
App-Type FE	•	•	•	•
App FE			•	•
App Controls			•	•
Observations	168	5,284,311	72	306,956
R-squared	0.916	0.980	0.918	0.935

Notes: The sample period in all columns is January 2014–April 2014. Data in Columns (1) and (3) consists of monthly observations at the app-type level. Data in Columns (2) and (4) consists of monthly observations at the app level. Columns (1) and (2) include all apps. Columns (3) and (4) include all app-types without pre-existing categories and other non-game and game app types with fewer than 20,000 apps in 2012. Outcomes for Columns (1)–(4) are the natural logarithms of downloads at each aggregation level. Additional app-level controls include average app ratings, a dummy for whether an app is free or paid, the price of an app if it is paid and app-age specific fixed effects. The variable “Games \times Post” is a dummy variable equal to 1 for games (or game app-types for even columns) during and after March 2014. The variable “Games \times Post \times No Pre-Existing” is a dummy variable equal to 1 during and after March 2014 only for games/app-types that did not have pre-existing categories before March 2014. Standard errors are clustered at the app-type level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

C.3 Downloads: March and April 2014 Estimates

In the main text, I use January and February 2014 as the “Pre”-policy period for the difference-in-differences download regressions in Tables 3 and C2. March and

¹²Apps without pre-existing categories have fewer apps on average than game types with pre-existing categories or non-game types. I exclude the largest non-game and game types (by the mean number of apps in 2013) to address this concern.

April are the “Post”-policy period. I do this because the policy change happened in the middle of March 2014, so data from that month may not fully reflect the policy change.

Table C3 replicates key regressions from Tables 3 and C2 from the main text using only data from February and March 2014. The “Pre” policy period is February 2014 and the “Post” policy period is March 2014. Results are qualitatively similar but quantitatively smaller than in the main text.

Table C3: Downloads Difference in Differences: February and March Data

Outcome Variable:	(1) ln(Tot Type Dwnlds)	(2) ln(Downloads)	(3) ln(Tot Type Dwnlds)	(4) ln(Downloads)	(5) ln(Tot Type Dwnlds)	(6) ln(Downloads)	(7) Post/Pre Δ ln(Downloads)
Games \times Post	1.081*** (0.220)	0.286** (0.132)	0.251** (0.113)	0.084* (0.050)	0.144*** (0.012)	0.072*** (0.009)	
Games \times Post \times No Pre-Existing			1.494*** (0.242)	1.791*** (0.082)			
Games \times Post \times Small Type					0.632*** (0.010)	0.921*** (0.119)	
Post/Pre Δ ln(N Apps)							-0.594*** (0.032)
Unit of Observation	App-Type	App	App-Type	App	App-Type	App	App
Sample Period	Feb 14/Mar 14	Feb 14/Mar 14	Feb 14/Mar 14	Feb 14/Mar 14	Feb 14/Mar 14	Feb 14/Mar 14	Feb 14/Mar 14
Sample	All	All	All	All	All Non-Games + Card, Casino, Arcade and Action	All Non-Games + Card, Casino, Arcade and Action	All Games
Observations	84	2,574,302	84	2,574,302	56	2,330,302	142,419
R-squared	0.833	0.981	0.944	0.989	0.998	0.992	0.868
Year/Month FE	•	•	•	•	•	•	
App-Type FE	•		•		•		
App FE		•		•		•	

Notes: The sample period in all columns is February and March 2014. Data in Columns (1), (3) and (5) consists of monthly observations at the app-type level. Data in Columns (2), (4), (6) and (7) consists of monthly observations at the app level. Columns (1)-(4) include all apps. Columns (5) and (6) include all non-game apps and Arcade, Action, Card and Casino game apps. Column (7) includes all game apps. Outcomes for Columns (1)-(6) are the natural logarithms of downloads at each aggregation level. The outcome for Column (7) is the difference between the natural log of app downloads in March 2014 and downloads in February 2014. Controls include year and month fixed effects and app-type or app fixed effects. Column (7) does not have fixed effects because it is a cross sectional regression in first-differences. Additional app-level controls include average app ratings, a dummy for whether an app is free or paid, the price of paid apps and app-age specific fixed effects. The variable “Games \times Post” is a dummy variable equal to 1 for games, or game app-types in March 2014. “No Pre-Existing” is a dummy variable equal to 1 for apps or app-types with no pre-existing categories before March 2014. “Small Split” is a dummy variable equal to 1 for Action and Casino games. “Post/Pre ln(N Apps in Category)” is the difference in the natural log of the number of apps in the category of app j in March 2014 and the number of apps in February 2014.

Standard errors are clustered at the app-type level for Columns (1)-(6) and are robust to heteroskedasticity in Column (7). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

C.4 Downloads: Placebos, Paid Apps Only and Alternative Outcomes

C.4.1 Downloads: Placebo Time Periods

Most of the main results in Section 4 are computed using a restricted data sample of four months, comparing January and February 2014 to March and April 2014. A possible concern may be that estimated effects are not caused by re-categorization

but by diverging time-trends between games and non-games or across game types. To test whether this is the case, I re-estimate the regressions using two comparable time periods without a re-categorization event: November 2013 to February 2014, and March 2014 to June 2014. For each sample, I estimate the effects of a non-existent re-categorization event: between December and January for the first sample, and between April and May for the second sample.

The first “placebo” sample verifies that download trends between games and non-games or across game types were not diverging before re-categorization took place. It also helps test whether it was the actual re-categorization or the *announcement* of re-categorization in December 2013 changed download outcomes. If changes in downloads were actually caused by re-categorization and improved consumer discovery, there is no reason to expect statistically significant differences in downloads before. The second “placebo” sample further verifies the effects of re-categorization. The policy was a permanent event - a consumer in May 2014 should have had as easy a time finding the “Music” game category as a consumer in April 2014. If re-categorization improved consumer discovery technology, these improvements should be locally persistent over time.¹³

I show results using alternative time periods in Table C4. These estimates replicate the main specifications shown in Tables 3 and C2.¹⁴ Panel (a) shows results using the Nov 2013 - Feb 2014 sample and panel (b) shows results using the March 2014 - June 2014 sample.

Nearly all estimates for the alternative time periods are statistically null. They are also generally substantially smaller in magnitude than estimates in the main text. There is some evidence of heterogeneous time-trends across game types without pre-existing categories and game types with pre-existing categories prior to re-categorization in Columns (3) and (4) in panel (a). However, relative to the baseline group of non-game types or apps, the total change in downloads for game app-types without pre-existing categories is still null. For both Columns (3) and (4), the sum of the “Games \times Placebo Post” and “Games \times Placebo Post \times No Pre-Existing” is not statistically significantly different from zero. These results show that differential changes in downloads between games and non-games occurred only during re-categorization. These results also show that the effects of re-categorization

¹³Over a longer period of time, changes in product assortment and entry may introduce additional congestion costs into the market, mitigating some immediate decreases in discovery costs. The changes in category informativeness, however, should be very persistent over time.

¹⁴Full time-varying estimates of treatment effects for specifications where I use data from January 2012 to December 2014 are in Figure C1. They also show the main download effects appear only following the actual re-categorization.

on downloads are persistent. Downloads four months after re-categorization were not statistically significantly different than the month after re-categorization. This suggests that the re-categorization event directly caused the change in downloads, consistent with permanent reductions in search costs for consumers due to an improvement in category informativeness and a reduction in the number of apps per category.

Table C4: Downloads Difference in Differences: Alternative Time Periods

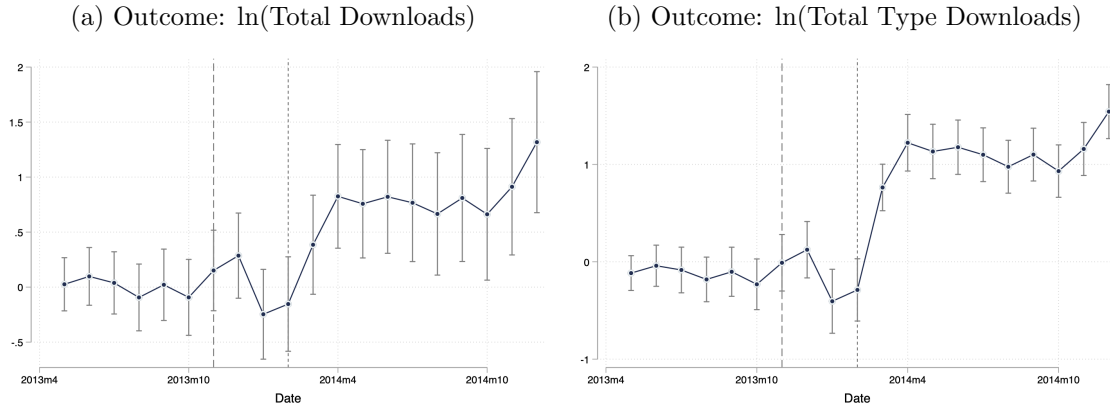
<i>Outcome Variable:</i>	(1) ln(Type Downloads)	(2) ln(Downloads)	(3) ln(Type Downloads)	(4) ln(Downloads)	(5) ln(Type Downloads)	(6) ln(Downloads)
Panel (a): Nov 2013 - Feb 2014						
Games \times Placebo Post	-0.131 (0.137)	-0.291 (0.199)	-0.219 (0.154) 0.158** (0.027)	-0.302 (0.198) 0.121** (0.035)	-0.210 (0.132)	-0.312 (0.149)
Games \times Placebo Post \times No Pre-Existing						
Games \times Placebo Post \times Small Type					0.019 (0.039)	0.105 (0.111)
Observations	168	4,892,297	168	4,892,297	112	4,427,584
R-squared	0.987	0.978	0.987	0.978	0.968	0.982
Panel (b): March 2014 - June 2014						
Games \times Placebo Post	0.223 (0.202)	0.065 (0.140)	0.110 (0.120) 0.204 (0.147)	0.056 (0.117) 0.064 (0.168)	0.124 (0.107)	0.079 (0.106)
Games \times Placebo Post \times No Pre-Existing						
Games \times Placebo Post \times Small Type					0.082 (0.074)	-0.030 (0.119)
Observations	168	5,496,539	168	5,496,539	112	4,935,476
R-squared	0.900	0.985	0.909	0.985	0.955	0.986
Unit of Observation: Sample	App-Type All	App All	App-Type All	App All	App-Type All Non-Games + Card, Casino, Arcade and Action	App All Non-Games + Card, Casino, Arcade and Action
Year/Month FE	•	•	•	•	•	•
App-Type FE	•		•		•	
App FE		•		•		•

Notes: The sample period in panel (a) covers November 2013 to February 2014. The sample period in panel (b) covers March 2014 to June 2014. Sample in Columns (1), (3) and (5) consists of monthly observations at the app-type level. Sample in Columns (2), (4) and (6) consists of monthly observations at the app level. Outcomes are natural logarithms of downloads at each aggregation level. Controls include year and month fixed effects, app-type fixed effects, or app fixed effects, depending on the column. Additional app-level controls for Columns (2), (4) and (6) include average app ratings, a dummy for whether the app is free or paid, the price of the app if it is paid, and app age-specific fixed effects. The variable “Games \times Placebo Post” is a dummy variable equal to 1 for games, or game app-types during and after January 2014 for panel (a) and during and after May 2014 for panel (b). Standard errors are clustered at the app-type level. *** p<0.01, ** p<0.05, * p<0.1

C.4.2 Downloads: Paid Apps

Downloads of free apps may not be accurate proxies of consumer app usage, as consumers can easily install and uninstall such apps from their phones without fully inspecting or using them. This is not the case for paid apps that consumers are required to spend money on upfront. I re-estimate the main regressions from Tables

Figure C1: Downloads Timing Tests



Notes: Each panel shows estimates of coefficients τ_t from Equation 3 at different aggregation levels. Panel (a) is estimated at the game/non-game level and panels (b) is estimated at the app-type level. Data from January 2012 to December 2014 is used throughout. Additional controls in each regression include year/month fixed effects, game/non-game or app-type fixed effects, and game/non-game or app-type specific trends. Standard errors for panel (a) are robust to heteroskedasticity and standard errors for panel (b) are clustered at the app-type level. 95% confidence intervals shown. In each panel, the first dashed vertical line represents the announcement of re-categorization and the second dashed vertical line represents the start of the re-categorization period.

3 and C2 in the main text after restricting the sample to paid apps. I also test for placebo effects using alternative time periods, as in Table C4.

Estimates for the sample of paid apps are in Table C5. Odd columns show estimates from aggregated app-type regressions with app-type fixed effects. Even columns show estimates from app-level regressions and even columns are at the app level with app fixed effects. Panel (a) shows estimates of the regressions for the January 2014 - April 2014 period, testing for the effects of a re-categorization event in March. Panel (b) shows estimates using the November 2013 - February 2014 period, testing the effects of a non-existent re-categorization event between December and January. Panel (c) similarly shows estimates using the March 2014 - June 2014 period, with a non-existent re-categorization event between April and May.

In each panel, the first two columns show the baseline average effects. The next two columns test for heterogeneity across game types that had pre-existing categories before the policy and those that did not. Such heterogeneity identifies changes in discovery costs through increasing informativeness. The last two columns test for heterogeneity across Arcade, Action, Card and Casino game types, where Action and Casino were much smaller before re-categorization. Such heterogeneity should

identify changes in discovery costs through reducing the number of apps per category and congestion.

Results are consistent with those in the main text and the robustness checks above. Panel (a) shows that on average, downloads for paid games increased over non-games after re-categorization. Paid games belonging to types without pre-existing categories and games belonging to types with fewer apps are driving the main effects. Estimates are statistically significant at the 95% confidence level and are larger than in the main text. Results for the two “placebo” events, before and after the actual re-categorization, show statistically null effects. These estimates confirm that consumer discovery costs fell in response to re-categorization.

C.4.3 Downloads: Alternative Outcomes

I use three alternative outcome variables to test the robustness of estimates in Section 4. Two of the alternative outcomes do not rely on the procedure described in Appendix B.2. As discussed by [Liebowitz and Zentner \(2020\)](#), the parametric assumptions used to generate most monthly download values in Appendix B.2 can produce biased estimates of actual downloads.

The first alternative outcome restricts the sample of apps to new apps: apps that entered the store in month t . For these apps, the approximation bias is minimal, since they are used to fit the model in Appendix B.2.

The second alternative outcome is a simpler proxy for monthly downloads: the difference in the number of user ratings for an app between two periods. For app j , downloads for period t are approximated by the number of user ratings in period t minus the number of user ratings in period $t - 1$ ($\text{N Ratings}_{jt} - \text{N Ratings}_{jt-1}$).¹⁵ This proxy relies on a simple, intuitive relationship - if a certain proportion of users who download an app also rate it, apps with more downloads will also have more ratings. Ratings on Google Play have to come from downloads, and it is unlikely that users will wait over a month to rate an app they downloaded. Such proxies have been used previously in papers studying mobile apps, such as [Kummer and Schulte \(2019\)](#). This approach has limitations, as the relationship between downloading and rating is not necessarily strictly monotonic in the number of downloads. Some types of apps may be very frequently downloaded but not frequently rated, whereas other apps are both frequently downloaded and rated. The proportion of users who rate apps may also decrease in app popularity. This would create a bias in the download

¹⁵On occasion, ratings and reviews disappear from the Google Play Store for various reasons including service term violations and the number of ratings falls between period $t - 1$ and period t . This occurs for less than 0.7% of observations. I bound changes in ratings to zero from below.

Table C5: Downloads Difference-in-Differences: Paid App Sample

<i>Outcome Variable:</i>	(1) ln(Type Dwnlds)	(2) ln(Dwnlds)	(3) ln(Type Dwnlds)	(4) ln(Dwnlds)	(5) ln(Type Dwnlds)	(6) ln(Dwnlds)
Panel (a): Jan 2014 - Apr 2014						
Games × Post	1.536*** (0.228)	0.489** (0.102)	0.594** (0.163)	0.252*** (0.040)	0.332** (0.069)	0.255*** (0.020)
Games × Post × No Pre-Existing			1.697** (0.315)	2.178*** (0.220)		
Games × Post × Small Type					1.322** (0.237)	1.560*** (0.154)
Observations	168	972,440	168	972,440	112	883,472
R-squared	0.761	0.978	0.876	0.986	0.953	0.990
Panel (b): Nov 2013 - Feb 2014						
Games × Placebo Post	0.052 (0.034)	-0.024 (0.057)	0.064 (0.043)	-0.029 (0.057)	-0.038 (0.056)	-0.022 (0.045)
Games × Placebo Post × No Pre-Existing			-0.022 (0.065)	0.049* (0.019)		
Games × Placebo Post × Small Type					0.314* (0.126)	0.020 (0.043)
Observations	168	942,428	168	942,428	112	858,336
R-squared	0.990	0.989	0.990	0.989	0.989	0.989
Panel (c): Mar 2014 - Jun 2014						
Games × Placebo Post	0.255 (0.138)	0.103 (0.069)	0.142 (0.084)	0.090 (0.072)	0.142 (0.084)	0.093 (0.052)
Games × Placebo Post × No Pre-Existing			0.203 (0.162)	0.096 (0.196)		
Games × Placebo Post × Small Type					0.104 (0.045)	0.023 (0.101)
Observations	168	958,186	168	958,186	112	866,545
R-squared	0.925	0.990	0.934	0.990	0.963	0.990
Unit of Observation: Sample	App-Type All Paid	App All Paid	App-Type All Paid	App All Paid	App-Type All Paid Non-Games + Paid Card, Casino, Arcade and Action	App All Paid Non-Games + Paid Card, Casino, Arcade and Action
Year/Month FE	•	•	•	•	•	•
App-Type FE	•		•		•	
App FE		•		•		•

Notes: The sample throughout all panels and columns only includes *paid* apps with non-zero prices. The sample period in panel (a) covers January 2014 to April 2014. The sample period in panel (b) covers November 2013 to February 2014. The sample period in panel (c) covers March 2014 to June 2014. Sample in Columns (1), (3) and (5) consists of monthly observations at the app-type level. Sample in Columns (2), (4) and (6) consists of monthly observations at the app level. Outcomes are natural logarithms of downloads at each aggregation level. Controls include year and month fixed effects, app-type fixed effects, or app fixed effects, depending on the column. Additional app-level controls for Columns (2), (4) and (6) include average app ratings, the price of the app, and app age-specific fixed effects. The variable “Games × Post” is a dummy variable equal to 1 for games, or game app-types during and after March 2014 for panel (a). The variable “Games × Placebo Post” is a dummy variable equal to 1 for games, or game app-types after during and after January 2014 for panel (b) and during and after May 2014 for panel (c). Standard errors are clustered at the app-type level. *** p<0.01, ** p<0.05, * p<0.1

proxy. For this reason, I choose to use downloads calculated according to B.2 as the main specification.

The last alternative outcome is the absolute number of predicted downloads rather than the natural log of downloads (using the predicted measure of downloads from Online Appendix B.2).

Results using these three outcomes are in Table C6. There are four panels in the

table. For ease of comparison, Panel (a) provides results using the baseline outcome from the main text. Panel (b) shows results using only new apps. Panel (c) shows results using the difference in the number of ratings to approximate downloads. Panel (d) shows results using the absolute number of predicted downloads. Odd columns aggregate data at the app-type level, and even columns use app level data. App level fixed effects are included for regressions in panels (a), (c) and (d). Panel (b) only includes app-type fixed effects, as I only observe each new app once. App level regressions in each panel are done using all non-game apps and game apps belonging to app-types without categories before the policies. I pick this sample as discovery costs should fall for this set of game apps (see panel (a) of Table C2). Columns (1) and (2) use the baseline January 2014 - April 2014 four-month sample period. Columns (3) and (4) use November 2013 to February 2014 as the sample period, with a “placebo” event between December 2013 and January 2014. Columns (5) and (6) use March 2014 to June 2014 as the sample period, with a “placebo” event between April 2014 and May 2014.

Results are qualitatively equivalent to the main specification. Results for ratings based downloads are different in magnitude than in the main text because of the different definition. However, downloads statistically significantly increase for games relative to non-games in March and April 2014 relative to January and February. The same does not happen in May and June 2014 relative to March and April, or in January and February 2014 relative to November and December 2013. Estimates in panel (b) using the sample of new apps are also similar to those in the main text.

Table C6: Downloads Difference in Differences Estimates: Alternative Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
Panel (a): Baseline						
<i>Outcome Variable:</i>	ln(Tot. Type Downloads)	ln(App Downloads)	ln(Tot. Type Downloads)	ln(App Downloads)	ln(Tot. Type Downloads)	ln(App Downloads)
Games × Post	1.399** (0.288)	2.214*** (0.302)				
Games × Placebo Post			-0.131 (0.137)	-0.174 (0.173)		
Games × Placebo Post					0.223 (0.202)	0.120 (0.273)
Observations	168	4,646,394	168	4,294,910	168	4,825,839
R-squared	0.780	0.982	0.987	0.982	0.900	0.986
Panel (b): New App Sample						
<i>Outcome Variable:</i>	ln(Tot. New Type Downloads)	ln(New App Downloads)	ln(Tot. New Type Downloads)	ln(New App Downloads)	ln(Tot. New Type Downloads)	ln(New App Downloads)
Games × Post	1.840** (0.318)	2.440*** (0.128)				
Games × Placebo Post			0.486 (0.207)	-0.315 (0.172)		
Games × Placebo Post					-0.045 (0.085)	-0.087 (0.120)
Observations	168	378,987	168	481,961	168	363,619
R-squared	0.821	0.622	0.849	0.507	0.931	0.706
Panel (c): Ratings Based Downloads						
<i>Outcome Variable:</i>	ln(Tot. Type Δ Ratings)	ln(App Δ Ratings)	ln(Tot. Type Δ Ratings)	ln(App Δ Ratings)	ln(Tot. Type Δ Ratings)	ln(App Δ Ratings)
Games × Post	0.092** (0.025)	0.196* (0.062)				
Games × Placebo Post			0.068 (0.032)	0.172 (0.147)		
Games × Placebo Post					0.072 (0.039)	0.050 (0.062)
Observations	168	4,646,680	168	4,284,464	168	4,829,472
R-squared	0.992	0.918	0.987	0.877	0.991	0.917
Panel (d): Absolute Downloads						
<i>Outcome Variable:</i>	Tot. Type Downloads	App Downloads	Tot. Type Downloads	App Downloads	Tot. Type Downloads	App Downloads
Games × Post	1301604.410* (459,123.970)	739.440** (227.632)				
Games × Placebo Post			-217,949,749 (313,913.912)	-198.974 (106.365)		
Games × Placebo Post					417,688.494 (383,730.890)	110.491 (184.156)
Observations	168	4,646,394	168	4,294,910	168	4,825,839
R-squared	0.854	0.780	0.843	0.817	0.908	0.931
Unit of Observation:	App-Type	App	App-Type	App	App-Type	App
Sample Period:	Jan 14/Apr 14	Jan 14/Apr 14	Nov 13/Feb 14	Nov 13/Feb 14	Mar 14/Jun 14	Mar 14/Jun 14
Sample	All	All Non-Games + Games w/o Pre-Exist. Cats.	All	All Non-Games + Games w/o Pre-Exist. Cats.	All	All Non-Games + Games w/o Pre-Exist. Cats.
Year/Month FE	•	•	•	•	•	•
App-Type FE	•	•	•	•	•	•
App FE		•		•		•

Notes: Sample for odd columns includes all apps and for even columns includes all non-games and Adventure, Board, Education, Family, Music, Role Playing, Simulation, Strategy, Trivia and Word games. Sample period in Cols (1)-(2) covers Jan/Apr 2014. Sample period in Cols (3)-(4) covers Nov 2013 - Feb 2014. Sample period in Cols (5)-(6) covers Mar/June 2014. Odd columns sample Outcomes are defined as the title of each column/panel combination. Controls include year and month fixed effects, app-type fixed effects, or app fixed effects, depending on the column. Additional app-level controls for even columns include average app ratings, a dummy for whether the app is free or paid, the price of the app if it is paid, and app age-specific fixed effects. Panel (b) does not include app-level fixed effects. In Cols (1)-(2) “Games × Post” is a dummy variable equal to 1 for games, or game app-types during and after March 2014. In Cols (3)-(4) “Games × Placebo Post” is a dummy variable equal to 1 for games, or game app-types during and after January 2014. In Cols (5)-(6) “Games × Placebo Post” is a dummy variable equal to 1 for games, or game app-types during and after May 2014. Standard errors are clustered at the app-type level. *** p<0.01, ** p<0.05, * p<0.1

C.5 Downloads: Changes in Number of Apps per Category

This section shows how, for a given app, the number of other apps in its category changes between November/December 2013 and January/February 2014, January/February 2014 and March/April 2014, and March/April 2014 and May/June 2014. I do this by first calculating, for each app, the average number of other apps in its category in each two-month period.¹⁶ Then I calculate, for each app, the difference between two successive periods.

The distribution of changes appears in Figure C2. This figure has three panels representing the three sets of changes I examine. The distribution of changes in panels (a) and (c) is entirely different from the distribution of changes in panel (b). In panels (a) and (c), the number of other apps in the category of an app increase. This is consistent with the general growth in the number of apps on Google Play over time. Panel (b) shows that between Jan/Feb and Mar/Apr 2014, all apps experienced a drop in the number of other apps in their category. The variation in this drop represents differences between what category the app belonged to before re-categorization and its category/app-type after re-categorization.

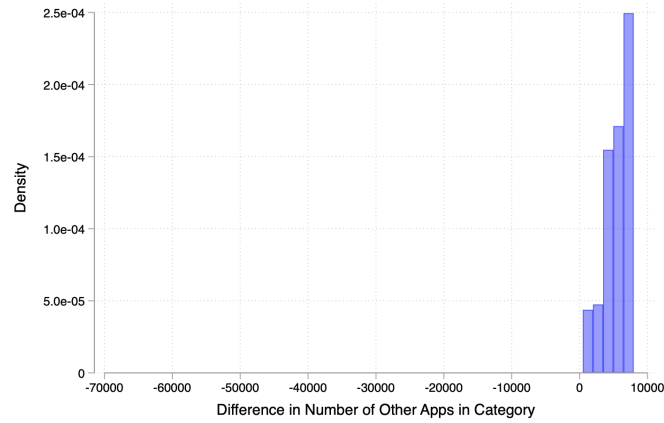
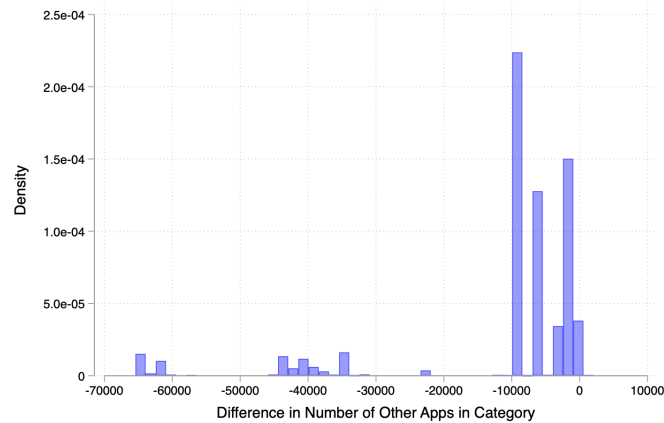
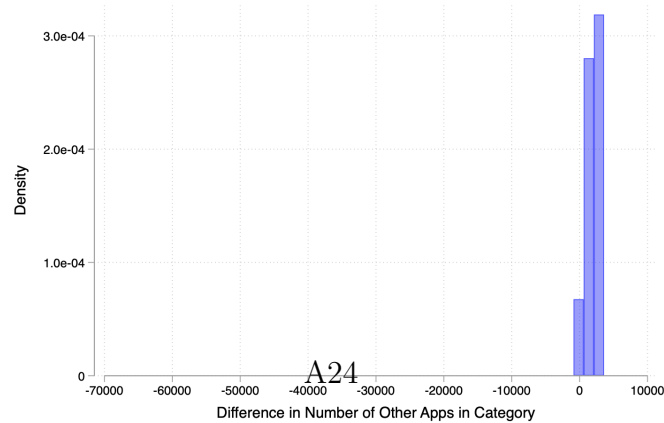
C.6 Downloads and Long Run Entry

In Column (3) of Table 3 in the main text, I show evidence of the effects of short-run changes in the number of apps in an app's category on app-level downloads. I do this using short-run changes induced by re-categorization, which move apps from broad categories to narrow categories that reflect their app types. Since where the apps end up is primarily determined by their pre-existing app-type and entry does not change much, this primarily reflects congestion rather than changes in the competition intensity (as reflected by the number of substitutes) for each app.

However, this does not necessarily mean that longer run changes in the number of apps of each app-type will have similar effects. As mentioned above, increases in the number of apps of each app-type could affect both competition intensity and congestion. I test for this in the data by relating long run differences in app-type entry to long run differences in downloads in the post re-categorization period. For a given app j , I calculate the difference in their downloads between December 2014 and March 2014. I then regress this difference on the difference in the number of apps of their type (which also coincides with their category) between December 2014

¹⁶I refer to categories here rather than app types since consumers use the stated category structure to search. See Section 2 for additional discussion. I use two-month periods, since these are comparable to the time periods I use in Table 3.

Figure C2

(a) $\ln(N \text{ Apps Jan/Feb 2014}) - \ln(N \text{ Apps Nov/Dec 2013})$ (b) $\ln(N \text{ Apps Mar/Apr 2014}) - \ln(N \text{ Apps Jan/Feb 2014})$ (c) $\ln(N \text{ Apps May/Jun 2014}) - \ln(N \text{ Apps Mar/Apr 2014})$ 

A24

Notes: Each panel shows the distribution of changes in app-level changes in the number of other apps in their category over time. For each app j , I calculate the difference in the natural log of the number of apps in their category between two successive periods. If app j is in category c^* in period t and category d^* in period $t + 1$, the difference is $\ln(N_{d^*,t+1}) - \ln(N_{c^*,t})$. In panel (a), the difference is between Jan/ Feb 2014 (on average) Nov/ Dec 2013. In panel (b), the difference is between Jan/Feb 2014 and Mar/Apr 2014. In panel (c), the difference is between Mar/Apr 2014 and May/Jun 2014.

and March 2014. The estimating equation is as follows:

$$\ln(Downloads_{j,Dec}) - \ln(Downloads_{j,Mar}) = \alpha(\ln(NApps)_{j,Dec}) - \ln(NApps)_{j,Mar}) + \beta X_j + \epsilon_j \quad (2)$$

where X_j are app characteristics I control for to account for unobservable heterogeneity not fully absorbed by the within-app differencing.

Estimates of this regression are in Table C7. The coefficient on the difference in the number of apps in the category is -0.66, suggesting that a one percent increase in the number of apps reduces app downloads by 0.66%.

Table C7: Long Run Changes in Downloads and Entry

Outcome:	(1) Dec/Mar $\Delta \ln(\text{Downloads})$
Dec/Mar $\Delta \ln(N \text{ Apps})$	-0.655*** (0.016)
Unit of Observation	App
Sample	All Games
Sample Period	Mar and Dec 2014
App Controls	•
Observations	121,134
R-squared	0.556

Notes: Sample includes all apps present in both March and December 2014. App-level controls include average app ratings, a dummy for whether the app is free or paid, the price of the app if it is paid, and app age-specific fixed effects. Standard errors are robust to heteroskedasticity. *** p<0.01, ** p<0.05, * p<0.1

These long-run estimates of the elasticity between changes in the number of apps in a category and app-level downloads are strikingly similar to short-run estimates in the main text. The coefficient on the short run re-categorization driven changes in the number of apps on the number of downloads is -0.65. This is reassuring, since the short-run re-categorization effect on congestion I identify in the main text seems to also be operating in the longer run. If changes in longer run entry were also generating pressure on app downloads through competition from additional substitutes, I would expect the coefficient in the long run regression to be substantially larger (in absolute terms). One possibility is that competition is already intense in the app market in March 2014 such that additional entry between March and December does not significantly increase it.

The OLS regressions estimated in this section may be subject to endogeneity concerns, as both downloads and app-type entry could be determined by common unobservable shocks. However, in many ways, the evolution of app-type entry after re-categorization is driven by changes caused by re-categorization. In Section C.7.1 I show that changes in entry after re-categorization are driven by whether the app-type's discovery costs fell during re-categorization. This means that app entry in

November 2014 (e.g., between November and December) was largely driven by the re-categorization which happened eight months prior.

C.7 Entry

C.7.1 Entry: Mechanisms

Sections 4 and Appendix C.2 show that re-categorization produced two main demand-side effects: an increase in the informativeness of categories in the store, and a reduction in congestion. Both of these improved consumers' ability to effectively browse the store, but some app-types (and ex-post categories) were more affected than others. The supply-side entry effects shown in Section 5.3.1 should be driven by these demand-side mechanisms. In that case, entry effects should display similar heterogeneity to download effects. For example, consumers became more informed about app-types that did not feature in the pre-policy categories. Entry in these app-types should increase more as well. I test for such heterogeneity in Table C8. As in Table C2, Column (1) tests for the effects of changes in the informativeness of categories by comparing app-types with and without pre-existing categories. As in Table 3, Column (2) tests for the effects of changes in congestion costs by comparing small and large app-types among those split from two pre-policy game categories. Small app-types had fewer apps before re-categorization, and they experience greater decreases in congestion after re-categorization.

Estimates confirm that average effects in Table 5 were primarily driven by app-types with greater changes in discovery costs. Entry increased more for game app-types that did not have pre-existing categories, as compared to game app-types with pre-existing categories (relative to non-game types). Among the app-types split off from pre-policy game categories, smaller app-types also had bigger changes in entry. App-types whose discovery costs were less affected by the policy have statistically null changes for both of the main supply-side outcomes. These results are also robust to alternative outcomes, such as the absolute number of entrants (Table C9).

Results in Tables 5 and C8 suggest that changes in consumer discovery costs are the main driving mechanism for product assortment changes in this market. The heterogeneity in entry effects also reflects theoretical predictions about the effects of discovery cost changes for different product types. Bar-Isaac et al. (2012) predict that "niche" products that benefit more from search cost reductions will experience the greatest increase in assortment. In this setting, the definition of "niche" products can include either app-types that had no pre-existing categories or "small types" that were relatively marginalized under the initial category structure.

Table C8: Entry Difference in Differences Estimates: Discovery Cost Channels

<i>Outcome Variable:</i>	ln(N Entrants) (1)	ln(N Entrants) (2)
Games \times Post	0.251* (0.137)	-0.086 (0.069)
Games \times Post \times No Pre-Existing	0.550*** (0.138)	
Games \times Post \times Small Type		0.833*** (0.202)
Unit of Observation	App-Type	App-Type
Time Period	Jan 12 / Dec 14	Jan 12 / Dec 14
Sample	All	All Non-Games + Action, Arcade Card and Casino
Year/Month FE	•	•
App-Type FE	•	•
Observations	1,470	980
R-squared	0.976	0.959

Notes: The sample period in all columns is January 2012-December 2014. Data in all columns includes monthly observations at the app-type level. Sample in Column (1) includes all game and non-game app-types. Sample in Column (2) includes all non-game app-types and Arcade, Action, Card and Casino game types. Outcomes in both columns are the natural log of the number of entrants in each app-type. Controls include year and month fixed effects, app-type fixed effects and app-type specific time trends. The variable “Games \times Post” is a dummy variable equal to 1 for games (or game app-types for even columns) during and after March 2014. The variable “Games \times Post \times No Pre-Existing” is equal to 1 during and after March 2014 only for app-types that did not have pre-existing categories before March 2014. The variable “Games \times Post \times Small Type” is a dummy equal to 1 during and after March 2014 only for Action and Casino game app-types. Standard errors in all columns are clustered at the app-type level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

C.7.2 Entry: Alternative Outcomes

Table C9 replicates Table 5 in the main text with alternative outcome variables. It uses *absolute* entry numbers. These regressions confirm the results from the log-transformed estimates in the main text. Absolute entry for the average app-type increases by about 1,500 apps, and for all games by almost 35,000 apps after re-categorization. As in the log estimates in Table 5, these are large treatment effects, given the size of the average game app-type. Columns (3) and (4) show that app-types where discovery costs fall by more in response to re-categorization are the ones driving the effects.

Table C9: Entry Difference in Differences Estimates with Alternative Outcomes

Panel (a): Absolute Number of Entrants				
Outcome:	N Entrants	N Type Entrants	N Type Entrants	N Type Entrants
Games \times Post	34,658.448** (16,259.260)	1,463.367*** (310.944)	1,134.432*** (394.645)	854.808** (348.410)
Games \times Post \times No Pre-Existing			592.083** (296.673)	
Games \times Post \times Small Type				1,214.406*** (438.110)
Observations	70	1,470	1,470	980
R-squared	0.876	0.775	0.776	0.776
Unit of Observation	Game/Non-Game	App-Type	App-Type	App-Type
Time Period	Jan 12 / Dec 14	Jan 12 / Dec 14	Jan 12 / Dec 14	Jan 12 / Dec 14
Sample	All	All	All	All Non-Games + Action, Arcade Card and Casino
Year/Month FE	•	•	•	•
App-Type FE				

Notes: Sample period in all columns and panels is January 2012 to December 2014. Sample in Column (1) includes monthly observations at the Game/Non-Game level. Sample in Columns (2) and (3) includes all game and non-game observations at the app-type level. Sample in Column (4) includes all non-game and Action, Arcade, Card and Casino app-type observations at the monthly level. Outcomes are the absolute number of new entrants at the game/non-game or app-type level. Outcomes in panel (b) are the average share of 1-star ratings at the game/non-game or app-type level. Controls include year/month fixed effects, a “Game” category group dummy for odd columns and app type fixed effects. Additional controls include game/non-game or app-type specific time trends. The variable “Games \times Post” is equal to 1 for games (or game types for even columns) during and after March 2014 and zero otherwise. Standard errors are robust to heteroskedasticity in Column (1) and clustered at the app-type level otherwise. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

C.7.3 Entry: Timing Tests

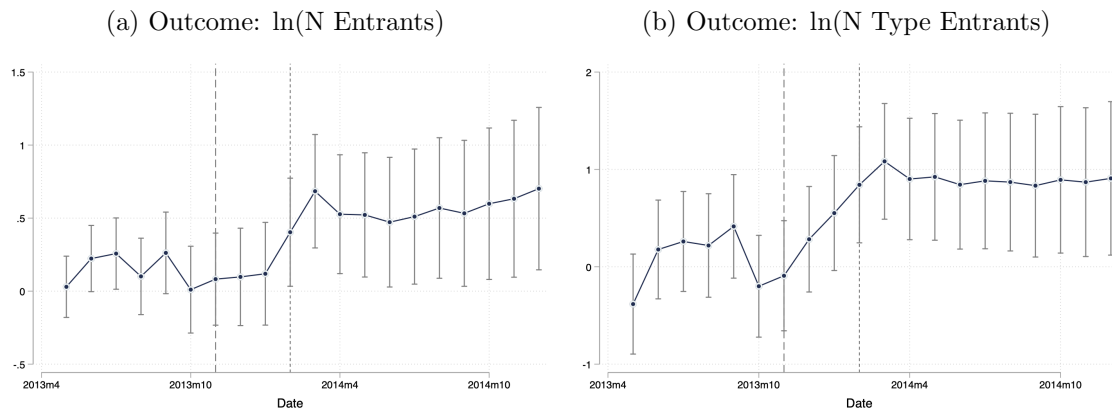
I allow treatment effects to vary over time by introducing interactions between monthly date dummies and the treatment group dummy. I estimate timing tests both at the aggregate game/non-game level and at the less aggregate app-type level. The estimating equation, for app-type c at time t is:

$$y_{ct} = \sum_{t=\text{re-cat. month}-10}^{\text{re-cat. month}+10} \tau_t (Game_c \times D_t) + \delta_c + \delta_t + \epsilon_{ct} \quad (3)$$

where y_{ct} is an outcome, and where τ_t s now capture period specific treatment effects relative to a baseline period. D_t is a dummy equal to one for observations during month t and zero otherwise. Since I have 10 periods after re-categorization, I test the 10 periods before re-categorization for parallel pre-trends relative to the time before April 2013. As before, I include game/non-game, app-type and time fixed effects, and game/non-game or app-type specific time trends. Figure C3 shows the period specific treatment effects for the main measure of entry used in the main text. For each outcome, game/non-game level results are on the left-hand panel, and app-type results are on the right-hand panel.

Entry estimates show that treatment effects become statistically significantly different from zero exactly around re-categorization. There are no treatment effect

Figure C3: Entry Timing Tests



Notes: Each panel shows estimates of coefficients τ_t from Equation 3 at different aggregation levels and with different outcomes. Panel (a) is estimated at the game/non-game level. Panel (b) is estimated at the app-type level. Data from January 2012 to December 2014 is used throughout. Additional controls in each regression include year/month fixed effects, game/non-game or app-type fixed effects, and game/non-game or app-type specific trends. Standard errors for panel (a) are robust to heteroskedasticity and standard errors for panel (b) are clustered at the app-type level. 95% confidence intervals shown. In each panel, the first dashed vertical line represents the announcement of re-categorization and the second dashed vertical line represents the start of the re-categorization period.

estimates which are statistically significantly different than zero (at the 95 percent confidence level) in the 10 periods before February 2014. February 2014 itself (two months following the announcement) has a statistically significant positive coefficient, likely representing a response by developers to the announcement of new categories in December 2013. Some apps may have entered the market early to position themselves in anticipation of the change.¹⁷ Entry response happens quickly after the announcement since apps have short development time. Developers can create simple apps in as little as a month.¹⁸ Point estimates are highest right after re-categorization takes place.

¹⁷The announcement did not set a strict date for the implementation of new categories, but said that the change will happen in the first quarter of 2014 (9to5Google.com)

¹⁸New entry could have come from multiple sources. Developers creating completely new apps, porting existing apps from the iOS store, or releasing already developed products into the market early.

C.8 Prices

Figure C4 plots three graphs showing price patterns in the Google Play Store. Panel (a) shows the ratio of mean paid game prices over mean paid non-game prices. Panel (b) shows the ratio of mean *new* paid game prices over mean *new* paid non-game prices. Prices do not appear to change substantially.

Panel (a) shows average prices for all games falling as compared to non-games, potentially due to increasing importance of in-app advertising and in-app purchases in the app economy. After re-categorization, average game prices increase and the ratio of game to non-game prices stabilizes. Lower discovery costs from the re-categorization could be the cause of the price changes.¹⁹ However, in absolute terms, the magnitudes of changes are small. In panel (b), it is apparent that there are no substantial differences in the prices of new games relative to non-games. The price ratio before and after is similar on average.

I estimate difference-in-differences regressions with both average prices and average entrant prices as outcome variables. Results for these regressions are in Table C10 at the game/non-game and at the app type. They show that there are no statistically significant differences between game and non-game paid app prices after re-categorization as compared to before. There is also no statistically significant heterogeneity (at the 95% confidence level) across game app-types that were more or less affected by re-categorization. This is true regardless of whether I look at all paid apps in panel (a) or only at new paid apps in panel (b).

Panel (c) of Figure C4 shows the ratio of the share of new paid games appearing in a given month (as a percentage of the total number of new games), over the share of new paid non-games (as a percentage of the total number of new non-games). Changes in the revenue streams of paid and non-paid apps (e.g., the increasing prevalence of in-app purchases) may result in changes in the number of entrants into the market. Such changes could also drive app entry and undermine the search mechanism explanation. This does not appear to be the case in the data. Panel (c) shows that there are no changes in the patterns of free and paid product entry between games and non-games after re-categorization.²⁰

In addition to the difference-in-differences estimates, I also test for whether changes in the number of other apps in a category affect a paid app's prices. As in Column (3) of Table 3 I use short run changes in the number of apps in game

¹⁹With lower costs, higher valuation consumers can discover more preferred game-apps more easily (Bar-Isaac et al. 2012).

²⁰There are substantial changes in the absolute share of paid apps that are entering into the market over time. The share of new paid products falls from over 30% to less than 10%. This pattern is consistent for both games and for non-game apps.

Table C10: Prices: Difference-in-Differences Estimates

Panel (a): All Paid Apps				
<i>Outcome Variable:</i>	Mean Price	Mean Price	Mean Price	Mean Price
Games \times Post	0.037 (0.036)	-0.037 (0.122)	-0.147 (0.152)	0.019 (0.105)
Games \times Post \times No Pre-Existing			0.197 (0.138)	
Games \times Post \times Small Type				-0.595* (0.347)
Games	9.966*** (0.758)			
Observations	70	1,470	1,470	980
R-squared	0.999	0.970	0.970	0.966
Panel (b): Paid Entrant Apps				
<i>Outcome Variable:</i>	Mean Price	Mean Price	Mean Price	Mean Price
Games \times Post	0.379 (0.443)	0.076 (0.423)	0.030 (0.470)	0.000 (0.459)
Games \times Post \times No Pre-Existing			0.082 (0.288)	
Games \times Post \times Small Type				-0.050 (0.910)
Games	28.088** (10.409)			
Observations	70	1,470	1,470	980
R-squared	0.999	0.970	0.970	0.966
Unit of Observation:	Game/Non-Game	App-Type	App-Type	App-Type
Sample:	All Paid	All Paid	All Paid	All Paid
Sample Period:	Jan 12/Dec 14	Jan 12/Dec 14	Jan 12/Dec 14	Jan 12/Dec 14
Year/Month FE	•	•	•	•
App-Type FE		•	•	•

Notes: Sample period in all columns covers January 2012-December 2014. Sample in Column (1) consists of monthly observations at the Game/Non-Game level. Sample in Columns (2)-(4) consists of monthly observations at the app-type level. Outcomes in panel (a) are average prices for all paid apps at each aggregation level. Outcomes in panel (b) are average prices for all paid new entrants at each aggregation level. Controls include year and month fixed effects, and game/non-game fixed effects or app-type fixed effects, depending on the column. Additional controls include game/non-game or app-type time trends. The variable “Games \times Post” is a dummy variable equal to 1 for games, or game app-types during and after March 2014. Standard errors are robust to heteroskedasticity in Column (1) and are clustered at the app-type level in the remaining columns. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

categories due to re-categorization. Apps move from being in large categories with many other apps of different types to smaller categories with other apps of their own type. One explanation for the findings in the main text showing that downloads for games with more apps in their category fall is that these apps face more competition from imperfect substitutes. If this is the case, there should also be a link between the number of apps in a category and prices.

In Table C11 I show the results of a regression relating pre/post re-categorization

differences in the number of apps in game categories to pre/post differences in individual app prices. Coefficient estimates are both small in absolute terms and are statistically null. There is no relationship between changes in the number of apps in a category and changes in app prices. This suggests that changes in the number of apps in a category does not affect competition. Instead, it primarily affects the market by reducing congestion.

Table C11: Price Changes in Response to Changes in Number of Apps in a Category

	(1) Post/Pre $\Delta \ln(\text{Price})$	(2) Post/Pre ΔPrice
Post/Pre $\Delta \ln(\text{N Apps in Category})$	-0.000 (0.002)	-0.005 (0.009)
Observations	21,749	21,749
R-squared	0.449	0.272
Unit of Observation:	App	App
Sample Period:	Jan 14 / Apr 14	Jan 14 / Apr 14
Sample:	All Paid Games	All Paid Games
App Controls	•	•

Notes: Sample period in both columns covers January 2014 to April 2014. Sample includes monthly observations of all paid game apps present from January 2014 to April 2014. Additional app-level controls include average app ratings and app age-specific fixed effects. The outcomes are differences between app-level average price in March and April 2014 and app-level average price in January and February 2014. $\Delta \ln(\text{N Apps in Category})$ is the difference in the natural log of the number of apps in the category of app j after re-categorization (March and April 2014) and the natural log of the number of apps in the category of app j before re-categorization (January and February 2014). Standard errors are robust to heteroskedasticity. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

C.9 Google Trends Evidence of Consumer Awareness of Android Games and Non-Games

I do not observe Google’s advertising for the Google Play app store. Instead, I use Google Trends search volumes to proxy consumer awareness for Android Games and Android Apps. Figure C5 shows the weekly Google Trends volumes from January 2012 to December 2015. The top two panels compare Google Trends for the “Android Games” and “Android Apps” search queries. The middle panels compare “Google Play Games” and “Google Play Apps.” The last two panels compare “Google Play Games” and “iOS Games.” In all cases Google trends are measured relative to the maximum search volume over the period. The figures on the left are absolute search trends numbers and the figures on the right are search trend ratios. Google Play/Android Games volumes are always the numerators in the ratios.

The figures all show that there is substantial variation in search query volumes over the sample period. For example, there is a spike in search queries around

Christmas/the New Year. In all three sets of comparisons there is no spike in Google Play/Android search queries around the period of the re-categorization of the store (solid vertical red line). There is also no change in the relative search query ratio around the period of the re-categorization. In the first two panels (Android Games vs. Android Apps), the “Google Play Games” search query is trending upwards relative to the “Google Play Apps” search query. There is no change in this trend around the re-categorization period.

C.10 Developer Switching Between Games and Non-Games

Figure C6 shows a ratio of the number of existing non-game developers²¹ who produce a game in period t over the number of existing game developers who produce a non-game in period t . The ratio is greater than 1: there are more non-game developers switching to producing games than the other way around. The figure also shows that there is an increase in the ratio around the period of the game re-categorization (from 2 to 2.6). This is potentially consistent with a resource allocation story whereby developers have a fixed budget and have to choose between producing games and non-games. However, within 2 months of the re-categorization, the ratio falls to pre re-categorization levels.

This suggests that developer switching after re-categorization is a short term response. By comparison, the entry effects of re-categorization are a long term phenomenon. Period-specific treatment effects captured in Figure C3 show that the increase in the number of games relative to non-games persists all the way to the end of the sample (9 months after re-categorization). The magnitude of the treatment coefficient in the last month of the sample in Figure C3 is as large as the magnitude in the second month after re-categorization. This suggests that the magnitude of the treatment effect cannot be explained by developers switching from producing non-games to producing games.

D Appendix D

D.1 Additional Demand Model Parameter Estimates

This section shows and discusses additional coefficient estimates of demand estimates from Column (4) of Table 4. Table D1 shows estimates for coefficients on lagged

²¹Defined as those who only developed non-games in the past.

downloads (q_{jt-1}), and various proxies for app quality - the number of screenshots, app size in MB, and whether or not an app has a video preview.

Table D1: **Additional Table 4 Column (4) Parameter Estimates**

γ Estimates	
ln(Lag App Downloads)	0.034*** (0.004)
ln(Size in MB)	0.046*** (0.006)
N Screenshots	0.008*** (0.001)
Video Preview Dummy	0.072*** (0.017)
Paid App Dummy	0.107 (0.194)

The coefficient on lagged downloads is positive. It suggests that apps with more past downloads are easier to find by consumers. It is also consistent with previous findings in the empirical literature on online product ranking-based discovery frictions (e.g., [Ursu 2018](#)). The positive coefficients on variables capturing app quality generally go in the expected direction. more screenshots, a video preview, and bigger app size should reflect higher app quality and generate additional consumer utility.

Figure [D1](#) plots estimates of app-type specific differences in pre/post re-categorization fixed effects. It shows substantial heterogeneity across app-types. On average, the ten app-types that did not have a pre-existing category (Adventure, Board, Educational, Family, Music, Role Playing, Simulation, Strategy, Trivia and Word) experience larger average increases in utility as compared to the eight app-types that had explicit categories before (Action, Arcade, Card, Casino, Casual, Puzzle, Racing and Sports).²² These effects are quantitatively large: on average, utility increases by 1 dollar for consumers from buying an app belonging to an app-type that did not have a category before the change, holding everything else constant.²³ This is consistent with reduced form evidence from Section [C.2](#), showing that re-categorization

²²The main exception for this group is Action games, which have a change in fixed effects comparable to some of the other app-types. This is possibly because it was grouped together with the Arcade app-type before re-categorization, leading to substantial improvement in the consumer search process.

²³Re-categorization also increases average utility for consumers from purchasing other app-types, although effects there are less than half the size on average (except for Action games, see previous footnote). This is likely because the informativeness changed for the other app-types as well, albeit

increased downloads for those app-types more than for the second group of app-types, and suggests that informativeness of the category structure increased after re-categorization.

In Appendix D.4 I show that this change is driven by the re-categorization, rather than some other average app-type time-varying differences.

D.2 Demand Model with Search

This section describes a demand model with search following the consideration set approach of Moraga-González et al. (2015).²⁴ Consumers choose a single product out of a set of N products. For each product j , consumers obtain utility $u_{ij} = \delta_j + \epsilon_{ij}$. Consumers are not fully informed about products: they do not know their ϵ_{ij} s. Search resolves this uncertainty. Consumers in this market first choose a consideration set A of products and pay a set-specific search cost. They find out the ϵ s of those products and pick a product j out of subset A . In this application, products in subset A can be located across multiple categories. Subsets are unobserved to the econometrician. Consumers know the expected utility (or inclusive value) they obtain from the products in subset A : \bar{U}_A .²⁵ Consumers incur subset-specific search costs (c_{iA}) such that the utility of consumer i choosing subset A is:

$$u_{iA} = \bar{U}_A - c_{iA} = \bar{U}_A - \left(\sum_{r \in A} \theta \psi_r + \lambda_{iA} \right) \quad (4)$$

where ψ_r reflects a deterministic “distance” between the consumer and product r in set A . λ_{iA} is a consumer/choice set specific search cost shock, which I assume is EV type 1 distributed mean zero with a standard error normalized to 1.²⁶ This shock can be interpreted as an information shock - word of mouth from friends or family. θ is effectively the average marginal search cost for consumers in the market. As with consumer utility, search costs have no unobservable heterogeneity aside from the idiosyncratic shock.

Due to the idiosyncratic error terms on both search costs and consumer utility, the unconditional probability of a consumer choosing product j is:

in relatively minor ways. For example, a consumer looking for a Card game may be less confused about what kinds of Card games are in the “Cards” category (i.e., no family card games, no music card games).

²⁴It is also similar to Goeree (2008) and Honka et al. (2017).

²⁵In a multinomial logit model, this is simply $\log[1 + \sum_{r \in A} \exp(\delta_r)]$. Consumers also always have the outside option, regardless of the set they consider.

²⁶ θ can vary across products or product groups.

$$P_j = \sum_{A \in A_j} P_A P_{j|A} \quad (5)$$

where A_j is the set of all subsets that product j belongs to, P_A is the probability of a consumer choosing subset A From the set of all possible subsets and $P_{j|A}$ is the probability that consumer i picks product j out of subset A . The unconditional probability P_j is equivalent to the observed market share of product j (s_j). [Moraga-González et al. \(2015\)](#) show that it is possible to “integrate out” the unobservable subsets and obtain the following closed form expression for s_j .²⁷

$$s_j = \frac{\frac{\exp(\delta_j)}{1 + \exp(\theta\psi_j)}}{1 + \sum_{k \in N} \frac{\exp(\delta_k)}{1 + \exp(\theta\psi_k)}} \quad (6)$$

where the denominator sums up over all products in the market (N) rather than over specific subsets. This expression is effectively the standard multinomial logit model except that the market share of product j is shaded down by how hard it is to find (ψ_j). I include the “discovery cost” variables from the model in the main text ($N_{c^*(j)}$ and R_{jc}) in ψ_j . The resulting expression is similar to the market share specification in Equation 4 in the main text. Setting $\exp(\gamma \ln(N_{c^*(j)}) + R_{jc}\kappa) = \frac{1}{1 + \exp(\theta\psi_j)}$ and introducing an additional nested logit error term equates the two.²⁸

I estimate this model using non-linear GMM with the same instruments used to estimate the linear demand model in the main text. Parameter estimates from this model are in Column (3) of Table D3. These are qualitatively similar to demand estimates from the main text. Note that signs for the “search cost” parameters are flipped relative to results in the main text because of how they enter into the model.

While this is a reasonable approach to modelling consumer product discovery and demand in the mobile app market, there are potentially many other ways in which consumers search the market. This model also does not easily allow controlling for additional unobservable heterogeneity with aggregate product level data. I choose to use the simpler linear demand model in the main text. It does not make specific assumptions about the consumer search process, but is broadly consistent with many predictions from search literature.

²⁷The assumption that the “distance” of products in a consideration set is additive in the set’s search costs is key for obtaining a closed form expression for choice probabilities.

²⁸This consideration-set based model allows for unobservable heterogeneity in consumer preferences, such as a consumer/category specific shock. However, the standard market-share inversion procedure for nested logit models does not apply to the consideration set model, and it would have to be estimated by simulation. I do not include additional unobservable heterogeneity for this reason.

D.3 Additional Demand Model Regressions

Table D2: 1st Stage Supporting Regression

Outcome Variable:	(1) ln(N Category Apps _t)
ln(N Category Apps _{t-1})	0.142*** (0.022)
Mean Category Rating _t	-0.279*** (0.065)
ln(Category Downloads _t)	0.387*** (0.034)
Category FE	•
Observations	624
R-squared	0.953

Notes: The sample includes monthly category-level observations from February 2012 to December 2014. Standard errors are robust to heteroskedasticity. *** p<0.01, ** p<0.05, * p<0.1.

D.4 Placebo Time-Varying Fixed Effects

In the main text, I include two sets of time-varying fixed app-type fixed effects: a set of app-type fixed effects that turns on before re-categorization takes place, and a set of app-type fixed effects that turns on after re-categorization takes place. The difference in these fixed effects is in Figure D1 in Appendix D.1. It shows app-type level welfare changes. I interpret these changes as being driven by re-categorization improving information, but they could also be caused by other time varying heterogeneity within app-types. For example, Educational games have the biggest fixed effect change, which could be the result of consumers liking educational games more over time.

To test whether the variation in fixed effects is driven by the re-categorization event, I introduce a specification of the model with three sets of time-varying app-type fixed effects. The first set of app-type fixed effects is active from March 2012 to February 2014. The second set of app-type fixed effects is active only during April 2014 (March 2014 is omitted from the data) and the last set of app-type fixed effects is active from May 2014 to December 2014. The change between the first two sets identifies information effects just around re-categorization. The change between the second two sets identifies whether there were other changes over time. If changes in app-type fixed effects primarily capture changes in category informativeness, I should

Table D3: Additional Demand Estimates

	(1)	(2)	(3)
γ		-0.392*** (0.026)	0.374*** (0.016)
$\gamma \times \text{New App}$		-0.012 (0.011)	
σ	1.404*** (0.069)	0.708*** (0.027)	
β_{price}	-3.800*** (0.119)	-0.833*** (0.111)	-1.470*** (0.041)
$\ln(\text{Lag App Downloads})$		0.036*** (0.004)	-0.282*** (0.008)
$\ln(\text{Size in MB})$	0.078*** (0.007)	0.046*** (0.006)	0.099*** (0.002)
Video Preview Dummy	-0.016 (0.020)	0.072*** (0.017)	0.138*** (0.005)
N Screenshots	0.000 (0.002)	0.008*** (0.001)	0.022*** (0.000)
Paid and New App Dummies	•	•	•
App Age FE	•	•	•
App Rating FE	•	•	•
Year/Month FE	•	•	•
Developer FE	•	•	•
App Type \times Pre/Post Re-Categorization FE	•	•	•
Observations	4,152,147	4,152,147	4,167,060

Notes: The sample includes monthly observations of all free and paid game apps in the Google Play Store from March 2012 to December 2014, excluding March 2014. Column (1) shows estimates of a nested logit model without discovery friction controls. Column (2) shows estimates of the model from the main text with heterogeneous discovery frictions for new and incumbent apps. Column (3) shows estimates of the search and demand model described in Online Appendix D.2. “App Rating FE” are a set of dummies representing the average rating of app j in period t within 0.5 stars. Apps with 2 stars or less are the “baseline” category for “App Rating FE.” *Year/MonthFE* include year and month dummies. Instruments for price and for σ include the ratings of other apps in the same category, the number of screenshots of other apps of the same app-type and the average size of other apps of the same app-type. Instruments for lagged downloads for app j include functions of further lags in app j downloads (2 and 3 periods before period t). The instrument for the number of apps in the category is described in the main text and is the residual of the regression in Table D2. Standard errors are clustered at the app level in Columns (1) and are (2) and robust to heteroskedasticity in Column (3). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

not see substantial differences between the April 2014 and May-December 2014 fixed effects where informativeness was constant.

Figure D2 shows both sets of differences for each app-type and their computed 95% confidence intervals. The results suggest that consumer utility from app types changed around the period of re-categorization and not after. Most of the “placebo” fixed effect differences are statistically zero at the 95% confidence level. Even for the

few app-types where these differences are not statistically zero, they are very small in magnitude relative to the true difference in fixed effects between the pre- and post-re-categorization.

D.5 Distributions of Main Model Welfare Effects

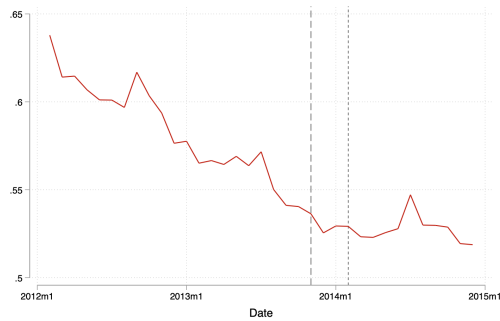
Figure [D3](#) shows the full distribution of welfare outcomes generated by the randomized simulations. The figures show that the randomizations do not substantially change the main effects.

Online Appendix References

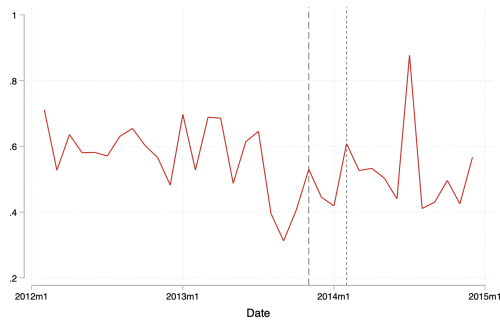
- Bar-Isaac, H., G. Caruana, and V. Cuñat Martínez (2012). Search, design and market structure. *American Economic Review* 102(2), 1140–1160.
- Chevalier, J. A. and D. Mayzlin (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research* 43(3), 345–354.
- Eeckhout, J. (2004). Gibrat’s law for (all) cities. *American Economic Review* 94(5), 1429–1451.
- Garg, R. and R. Telang (2013). Inferring app demand from publicly available data. *MIS Quarterly* 37(4), 1253–1264.
- Goeree, M. S. (2008). Limited information and advertising in the us personal computer industry. *Econometrica* 76(5), 1017–1074.
- Honka, E., A. Hortaçsu, and M. A. Vitorino (2017). Advertising, consumer awareness, and choice: Evidence from the us banking industry. *The RAND Journal of Economics* 48(3), 611–646.
- Kummer, M. and P. Schulte (2019). When private information settles the bill: Money and privacy in google’s market for smartphone applications. *Management Science* 65(8), 3470–3494.
- Leyden, B. T. (2018). There’s an app (update) for that. mimeo.
- Liu, Y., D. Nekipelov, and M. Park (2014). Timely versus quality innovation: The case of mobile applications on itunes and google play. *NBER Working Paper*.
- Moraga-González, J. L., Z. Sándor, and M. R. Wildenbeest (2015). Consumer search and prices in the automobile market.
- Ursu, R. M. (2018). The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions. *Marketing Science* 37(4), 530–552.

Figure C4: Prices

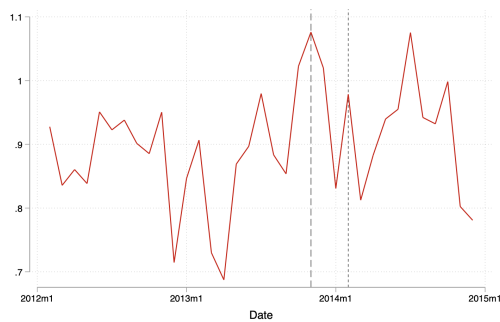
(a) Outcome: Mean Game Price / Mean Non-Game Price



(b) Outcome: Mean Entrant Game Price / Mean Entrant Non Game Price

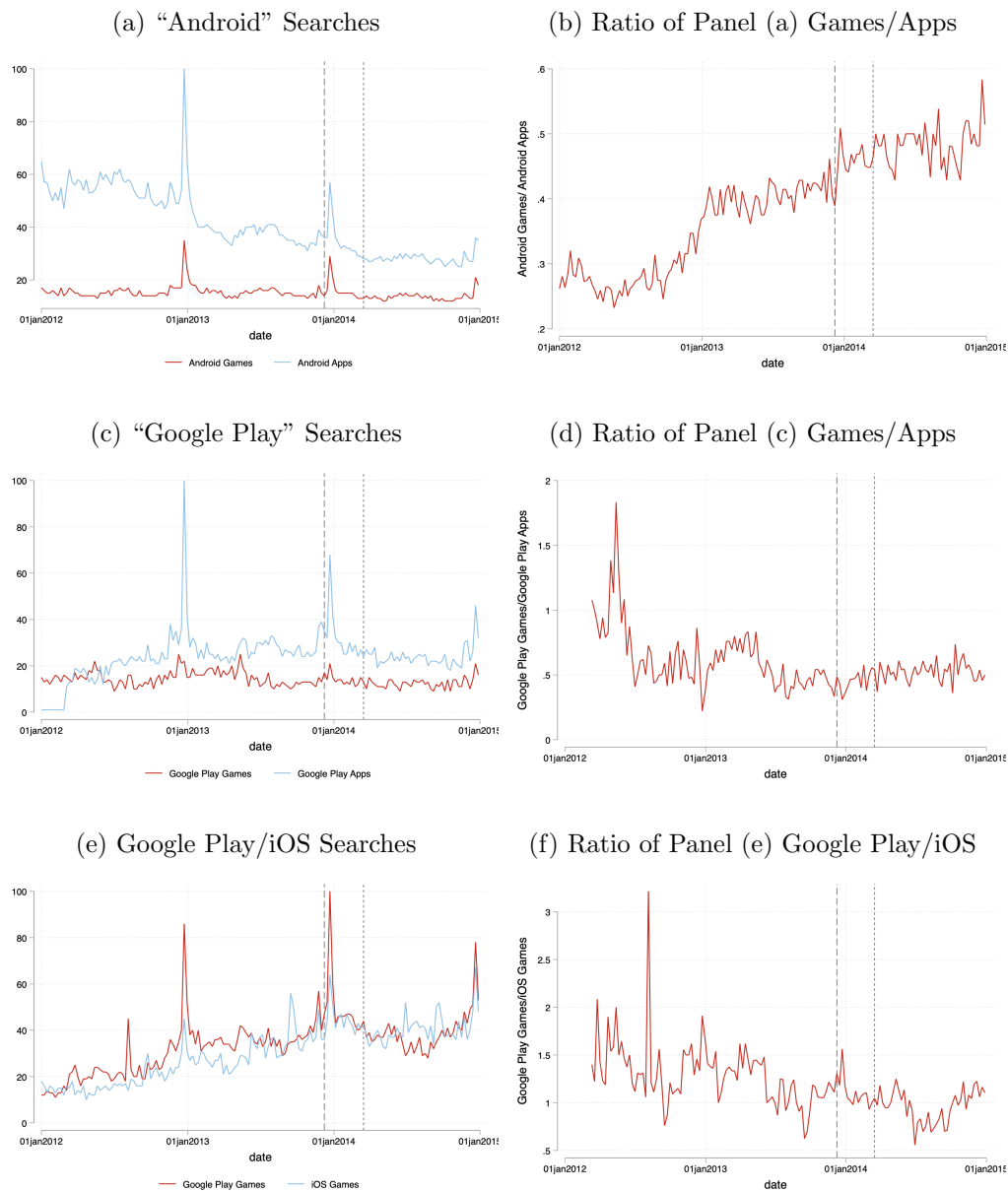


(c) Outcome: Share Paid Game Entrants / Share Paid Non-Game Entrants



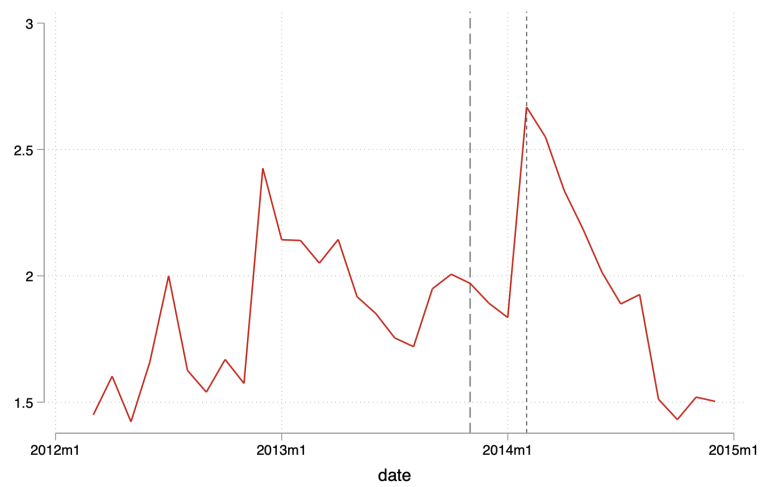
Notes: Panel (a) shows a ratio of mean monthly game app price over mean monthly non-game app price using all paid apps. Panel (b) shows a ratio of mean monthly game app price over mean monthly non-game apps price using only paid entrants. Panel (c) shows a ratio of the monthly percentage of new game apps that are paid over the monthly percentage of new non-game apps that are paid. In all panels, the first dashed vertical line represents the re-categorization announcement and the second dashed vertical line represents the start of the re-categorization period.

Figure C5: US Google Search Trends



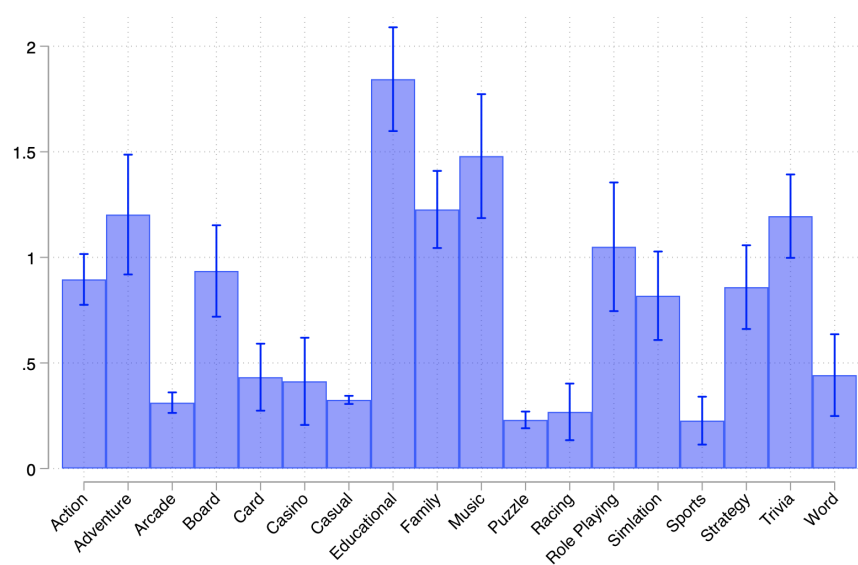
Notes: Panels (a), (c) and (e) show daily Google Trend search volume estimates for different queries. In each of the panels, numbers are normalized relative to maximum search volume which is set to 100. Panels (b), (d) and (f) show ratios of the numbers in panels (a), (c) and (e), respectively. In all panels, the first dashed vertical line represents the re-categorization announcement and the second dashed vertical line represents the start of the re-categorization period.

Figure C6: Ratio of Switching Developers: $\frac{\text{Non-Game to Game}}{\text{Game to Non-Game}}$



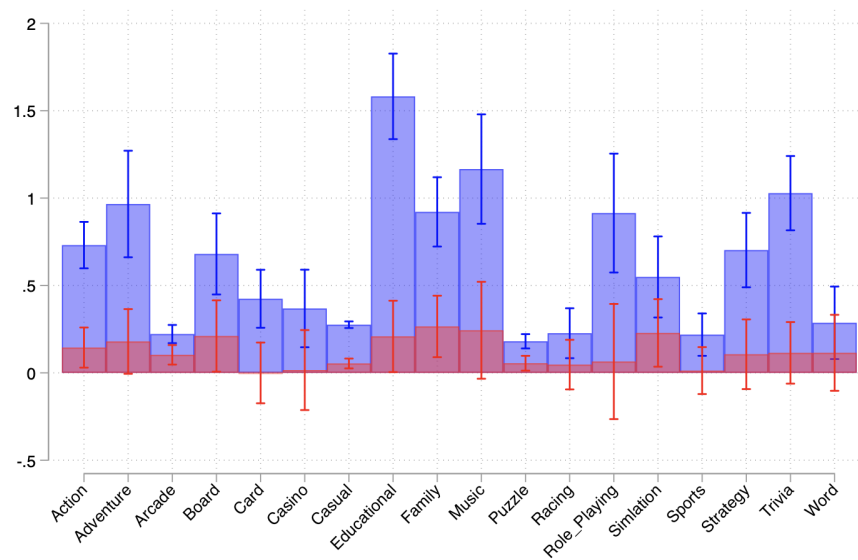
Notes: This figure shows a monthly ratio. In each month t the numerator is the number of developers who produced a non-game app in any period before t and produced a game app in period t . The denominator is the number of developers who produced a game app in any period before t and produced a non-game app in period t . The first dashed vertical line represents the re-categorization announcement and the second dashed vertical line represents the start of the re-categorization period

Figure D1: Difference in App-Type Fixed Effects



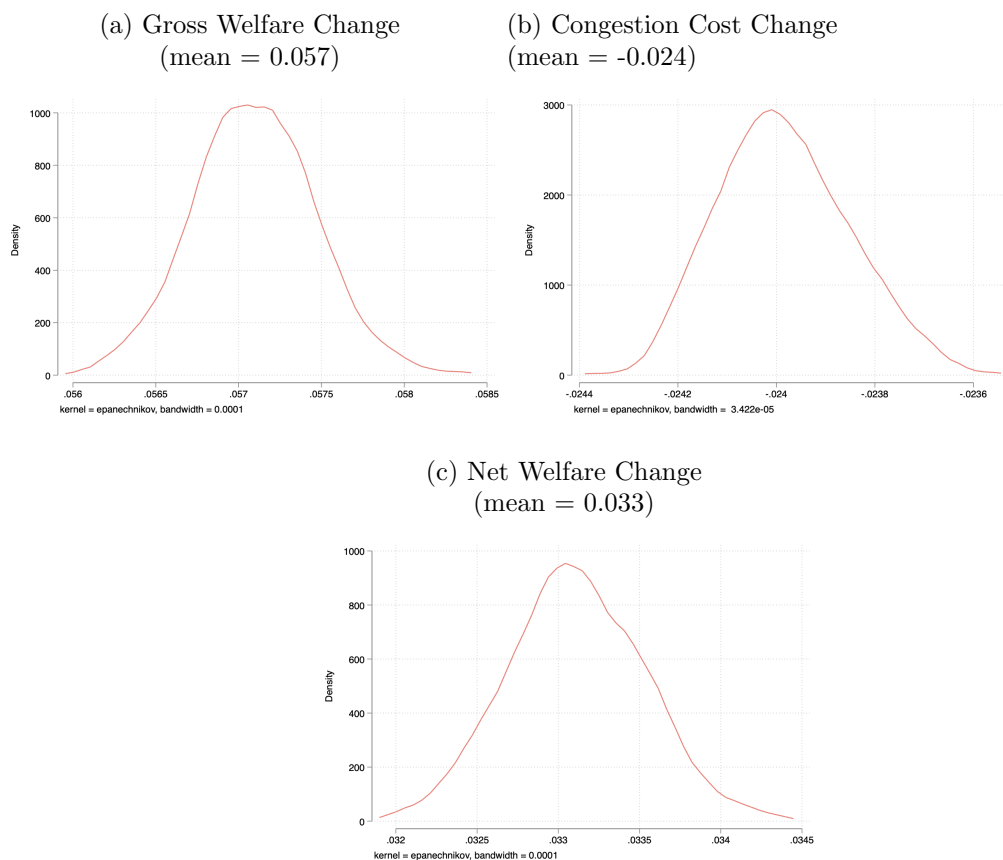
Notes: Each column shows the difference in estimated app-type fixed effects based on the model described in Section 5.1: the fixed effect for app-type c in the post- re-categorization period, minus the fixed effect value for app-type c in the pre- re-categorization period. 95% calculated confidence interval for this difference is shown.

Figure D2: Differences in App-Type Fixed Effects



Notes: Each column shows two sets of differences in estimated app-type fixed effects based on the model described in Section 5.1. The first difference, in blue, is between the app-type fixed effect for Mar 2012-Feb 2014 and the app-type fixed effect for April 2014. The second difference, in red, is between the app-type fixed effect for April 2014 and the app-type fixed effect for May 2014-December 2014. 95% calculated confidence intervals for both differences are shown.

Figure D3: Distribution of Welfare Effects Across Simulations



Notes: Panels show the distribution of welfare effects across 500 simulations.